

SV-F-1.1 poly 02
Texte du cours**Partie F : Génomique structurale et fonctionnelle**
Chapitre 1.1 : Génomes des cellules et des virus

La **génétiq**ue est la science de l'hérédité et de sa variation. Elle concerne tous les « organismes » possédant une information génétique transmissible de génération en génération : Les bactéries (et les archées) et les eucaryotes (= cellules), mais aussi les virus (acellulaires).

L'information génétique est portée par la molécule **d'ADN**, sous forme d'une séquence nucléotidique. Dans les cellules, l'ADN est ainsi le support permanent de l'information génétique.

La molécule **d'ARN** peut aussi porter une information génétique : Elle est transitoire dans les cellules, et peut correspondre au support principal dans le cas de certains virus.

La taille et l'organisation du génome peut varier selon les cellules ou virus considérés.

I. L'étude des génomes

A. Une diversité de génomes

Le **g**énome correspond à l'**ensemble de l'information génétique**.

Pour ce qui est de la nature du **s**upport **m**oléculaire, les génomes des cellules sont toujours constitués d'ADN double brin. Chez les eucaryotes, cet ADN double brin est linéaire, et constitue les chromosomes eucaryotes. Chez les bactéries, cet ADN est en général circulaire, constituant le chromosome bactérien.

La nature moléculaire du génome est très diversifiée chez les virus : ADN ou ARN, simple ou double brin, linéaire ou circulaire...

La **t**aille des génomes est très variable selon les organismes :

- Très petite taille chez les virus : de 5 kb à quelques dizaines de kb en général.
- Bactéries : en général de 600 kb à quelques Mb.
- Eucaryotes : le génome haploïde est d'une plus grande taille, pouvant aller d'une dizaine de Mb pour les « petits » génomes, jusqu'à plusieurs Gb dans le cas de « grands » génomes.

Au niveau de sa **l**ocalisation, le chromosome bactérien est cytoplasmique, regroupé au sein du nucléoïde.

Le génome eucaryote est essentiellement nucléaire, mais il est à noter que des petits génomes (sous forme d'ADN double brin circulaires) sont présents dans les organites semi-autonomes : mitochondries et chloroplastes.

On peut ainsi noter que les cellules possèdent un génome **p**ricipal (dans le nucléoïde ou le noyau) et parfois un génome **a**ccessoire (plasmides bactériens, organites semi-autonomes des eucaryotes).

Le génome viral a une localisation variable selon le type de virus et le moment du cycle de multiplication du virus : au sein de la particule virale, cytoplasmique, nucléaire.

B. Le séquençage des génomes

Le séquençage d'une molécule séquencée (ADN, ARN, protéine) correspond à la détermination de sa séquence : dans le cas d'un ADN ou d'un ARN (et donc d'un génome), il s'agit de déterminer la succession de nucléotides, et donc de bases azotées, qui constitue l'information génétique.

Ce séquençage repose sur diverses techniques moléculaires, dont la technique de Sanger (*cf. TP SV-F1*).

La majorité des techniques de séquençage (dont celle de Sanger, qui nécessite de partir d'une séquence connue, comme pour toutes les techniques utilisant une synthèse d'ADN par une ADN polymérase) ne permettent pas de séquences de longues molécules : On se limite souvent à l'obtention d'une séquence de moins de 250 paires de bases.

Le séquençage d'un génome (stratégie *shotgun*) nécessite donc de :

- 1- Fragmenter l'ADN génomique en de très nombreux fragments de petite taille ;
- 2- Puis de séquencer tous ces fragments ;
- 3- Et ensuite d'assembler les séquences obtenues, grâce aux séquences présentes dans deux fragments à la fois.

Ceci nécessite de séquencer une quantité totale d'ADN d'au minimum 10 fois la longueur du génome, afin d'obtenir un nombre de séquences permettant d'avoir assez de séquences chevauchantes et ainsi de reconstituer le génome.

Le développement des techniques de séquençage à haut débit, ainsi que l'apparition de techniques totalement différentes, permettant de séquencer des ADN de plusieurs centaines de kilobases, rendent le séquençage actuel des génomes peu coûteux et rapide, du fait de la possibilité d'avoir une profondeur de séquençage très importante.

Historiquement, le séquençage d'un génome débutait donc par la cartographie de grands fragments d'ADN les uns par rapport aux autres, puis chaque fragment était séquencé.

Afin de limiter la quantité de séquençage à réaliser, celui-ci peut être réalisé de proche en proche : Une première séquence est réalisée, puis on nouvelle amorce de séquençage est choisie, de manière à continuer le séquençage à partir de ce point.

C. L'interprétation des séquences génomiques

Une fois séquencé, le génome doit être annoté pour être exploitable. Cette annotation met en jeu des outils informatiques : On parle de **bioinformatique**.

D'une manière générale, l'annotation consiste en la recherche de séquences se rapprochant de séquences déjà connues.

On recherche ainsi, en particulier, les motifs caractéristiques des promoteurs et des séquences régulatrices.

La recherche des gènes passe par la détermination de **phases ouvertes de lecture**, ou **ORF** (*Open Reading Frame*) : Il s'agit de séquences d'une taille suffisante pour correspondre à un gène, débutant par un codon initiateur (ATG dans la majorité des cas) et se terminant par un codon STOP.

Ce type de technique est rendu difficile, chez les eucaryotes, par la présence de très nombreuses séquences non codantes, y compris au sein des gènes (introns en particulier)...

II. Le génome des bactéries

A. Un génome cytoplasmique

Les procaryotes ne présentent pas de membranes internes : leur ADN est donc cytoplasmique, même si le chromosome bactérien est « regroupé » au sein du nucléoïde, en position centrale dans la cellule bactérienne (on a donc une compartimentation fonctionnelle !).

1. Un chromosome circulaire

Le génome des procaryotes est essentiellement constitué d'un **chromosome circulaire**, de 600 kb à quelques mégabases, soit 1 000 fois moins d'ADN que les génomes eucaryotes, en moyenne.

E. coli : **4 600 kb, portant 4 300 gènes** (séquençage du génome entamé dans les années 1960, achevé en 1997).

Le génome des procaryotes est donc **haploïde** (sauf exception...).

2. Des plasmides accessoires

En plus du génome principal, les procaryotes possèdent un **génome accessoire**, sous la forme de molécules d'ADN qui se **répliquent de manière indépendante** : les **plasmides**. Ils ne sont **pas indispensables** à la survie de la cellule.

Leur taille peut être très variable, de 1 000 nucléotides à quelques centaines de kilobases. Ils sont le plus souvent circulaires, et portent des gènes divers qui ne sont pas essentiels lorsque la cellule se trouve dans un environnement favorable : ils confèrent un avantage sélectif dans certaines conditions ; ainsi, nombre d'entre eux codent des protéines de résistance aux antibiotiques (plasmides « R »).

Une cellule procaryote possède de 1-2 copies jusqu'à 50 copies.

Certains plasmides peuvent être **transférés** d'une bactérie à une autre dans certaines conditions. Il s'agit de la conjugaison bactérienne, qui permet un transfert horizontal d'information génétique. Le plasmide « F » peut être transféré d'une bactérie F⁺ à une bactérie F⁻.

B. Un génome compact

On peut remarquer que le génome des eubactéries est plus petit que celui des eucaryotes.

Mais la comparaison entre génome ne se limite pas à leur taille : elle peut se réaliser au niveau de la nature même du **contenu informatif** des génomes.

1. Une majorité de séquences codantes

Les génomes procaryotes ont une **fraction codante** importante, bien plus en moyenne que chez les eucaryotes, de l'ordre de **90%**.

Le génome bactérien est donc essentiellement constitué de gènes, avec très peu de séquences intergéniques.

Cette fraction codante correspond aux séquences transcrites et traduites en protéines. Toutefois un génome ne peut pas se limiter à ces seules séquences : des séquences régulatrices sont nécessaires au contrôle de l'expression génétique, aussi bien d'un point de vue purement « fonctionnel » (repérage des sites d'initiation de la transcription, etc.), comme d'un point de vue « adaptatif » (expression génétique adaptée aux besoins cellulaires).

Un génome s'exprime donc sous la forme d'un transcriptome (ensemble des informations transcrites) et d'un protéome (ensemble des protéines présentes).

2. Les opérons bactériens

a. Principe général d'un opéron

De nombreux gènes eubactériens sont regroupés en structures organisées sous le contrôle d'une même séquence régulatrice : on parle alors d'**opéron**.

Un **opéron** est une **unité transcriptionnelle** : un unique promoteur régule la transcription d'un ARNm (dit polycistronique), qui porte les séquences codant pour plusieurs protéines. Il s'agit de protéines intervenant dans une **même voie métabolique**, un **même phénomène** cellulaire.

E. coli : environ un quart des gènes sont organisés en opéron. Cette proportion est variable selon les espèces.

b. L'opéron lactose

L'**opéron lactose** est un exemple classique d'opéron bactérien. L'élucidation de son fonctionnement fait partie des travaux qui ont permis à François Jacob, Jacques Monod et André Lwoff d'obtenir le Prix Nobel de médecine en 1965.

Des bactéries, telles que *E. coli* utilisent habituellement le glucose comme source de carbone et d'énergie. Cependant lorsque cet ose est rare, *E. coli* peut utiliser le lactose comme source de carbone, même si ce dissaccharide ne se situe sur aucune des voies métaboliques principales.

Le métabolisme du lactose fait intervenir plusieurs gènes, regroupés dans le génome d'*E. coli* au sein d'un même opéron, sous le contrôle d'une même séquence régulatrice : l'**opéron lactose**.

L'opéron lactose comprend trois gènes, dits gènes de structure, codant des enzymes :

- Gène **lacZ** : **β-galactosidase**
- Gène **lacY** : **perméase** (galactoside perméase), nécessaire au transport du lactose dans la cellule
- Gènes **lacA** : **transacétylase**, dont le rôle reste inconnu à l'heure actuelle...

Ces trois protéines sont synthétisées à partir d'un unique ARNm (on parle d'ARNm polycistronique).

Un autre gène, le gène **lacI** (inhibiteur), est situé plus loin dans le génome : Il est exprimé de manière constitutive, et permet la synthèse d'une protéine tétramérique, le répresseur.

Le répresseur est une protéine capable de se fixer sur l'ADN, au niveau d'une séquence spécifique : l'**opérateur**. Cette fixation empêche l'ADN polymérase de se fixer sur le promoteur et de transcrire les gènes de structure : Il n'y a pas d'expression de l'opéron.

En présence de **lactose** dans le milieu, une très faible proportion de celui-ci pénètre dans la cellule, sous forme d'allolactose (un isomère du lactose). Cet allolactose est un ligand du répresseur. La fixation de l'allolactose sur le répresseur provoque un changement de conformation de ce dernier, qui est alors inactivé : Il ne peut plus se fixer sur l'ADN. L'inhibition de l'opéron lactose est ainsi levé, et les gènes de structure sont transcrits puis traduits.

III. Le génome des eucaryotes

Les eucaryotes sont des cellules compartimentées, qui possèdent un génome nucléaire et un génome extranucléaire dans les organites semi-autonomes.

A. Différentes localisations cellulaires

1. Un génome nucléaire de grande taille et fragmenté

Le génome des eucaryotes est plus grand que celui des procaryotes : d'une dizaine de Mb à plus de 100 Gb. Il est formé de molécules d'ADN **linéaires**, qui constituent les **chromosomes**, de tailles et de structure variable selon les espèces.

De plus, les génomes peuvent être **haploïdes** (chromosomes en un seul exemplaire) ou **diploïdes** (chromosomes en deux exemplaires) en fonction de la place de l'organisme considéré dans le cycle de développement de l'espèce.

En fonction de la place de la cellule dans le cycle cellulaire, un chromosome est formé d'une molécule d'ADN (**monochromatidien**) ou de deux molécules identiques aux erreurs de réplication près (**bichromatidien**).

Le génome humain haploïde est ainsi de **3.10⁹ pb** (3 Gb), réparti sur 23 paires de chromosomes ; chaque chromosome compte de 50 à 250 Mb.

Chez les organismes diploïdes, un chromosome de chaque paire est hérité du père, et l'autre est hérité de la mère. Génotype noté en indiquant les allèles présents sur chaque chromosome : AB/ab, Ab/aB par exemple.

2. Un génome extranucléaire proche des eubactéries

Les **organites semi-autonomes** (mitochondries et chloroplastes) des eucaryotes possèdent un génome propre. Il s'agit d'un petit génome d'ADN circulaire (en général), présent en plusieurs dizaines d'exemplaires par organelle.

La comparaison de séquences d'ARNr bactériens, eucaryotes nucléaires, mitochondriaux et chloroplastiques montre une forte proximité entre les génomes bactériens et ceux des organites semi-autonomes.

De plus les séquences régulatrices de ces génomes sont similaires à celles des eubactéries (et non des eucaryotes), leur génome est circulaire et non linéaire... La membrane interne des mitochondries et chloroplastes présente, enfin, une composition lipidique similaire à celle de membranes bactériennes.

Tous ces arguments suggèrent une origine bactérienne des organites semi-autonomes : c'est la **théorie endosymbiotique**.

B. Un génome nucléaire condensé

L'ADN nucléaire n'est pas « nu », mais associé à des protéines : il est présent sous forme de **chromatine**.

Les protéines associées à l'ADN sont de deux types : les **histones** et les protéines **non-histones**.

On peut distinguer plusieurs niveaux de condensation de l'ADN :

- La **fibres de 11 nm** : Correspond à l'association entre ADN et histones, formant des nucléosomes espacés d'environ 200 pb (dont 146 pb enroulées autour du nucléosome). Cette fibre correspond à l'**euchromatine sensu stricto**.
- La **fibres de 30 nm** : Regroupement des nucléosomes grâce aux histones H1.

On trouve la fibre de 30 nm dans l'**euchromatine** : l'euchromatine est en effet constituée pour les 10% (environ) des gènes les plus actifs de fibre de 11 nm, et de la fibre de 30 nm pour les 90% restant. L'**hétérochromatine** est constituée de fibre de 30 nm, avec un repliement supplémentaire rendant l'ADN inaccessible.

Les gènes présents dans l'euchromatine peuvent être transcrits, et donc exprimés (*mais cette expression n'est pas obligatoire*). Les gènes présents dans l'hétérochromatine ne sont pas exprimés. On peut distinguer :

- L'hétérochromatine **facultative** : Elle peut passer sous forme décondensée, et ainsi permettre l'expression génétique.
- L'hétérochromatine **constitutive** : Elle est en permanence condensée. On y trouve par exemple des gènes utilisés pendant le développement embryonnaire uniquement.
- La **fibres de 300 nm** : la présence de séquences SAR permet le repliement de la chromatine sur un axe protéique non-histone, en boucles de 4,5 à 112 kb.

Pendant les divisions cellulaires la chromatine connaît un état de compaction supplémentaire avec la mise en place des chromosomes mitotiques (méiotiques), grâce à l'intervention de diverses protéines non-histones.

On peut noter que les procaryotes présentent aussi une condensation de l'ADN, faisant intervenir des protéines avec un rôle analogue aux histones.

C. Un génome complexe

Le génome des eucaryotes est d'une taille plus importante que celui des procaryotes. Il présente une complexité importante en termes d'informations et de séquences présentes.

1. Un ADN non-codant abondant

Le génome humain, de près de 3 milliards de pb compte environ 23 000 gènes (voire un peu moins). La quantité d'ADN totale du génome des eucaryotes semble donc « trop » importante par rapport aux gènes qu'il contient. Une part importante du génome des eucaryotes est donc constituée d'autre chose que des gènes : on parle d'ADN non codant. Homme : **moins de 2% d'ADN codant** (exons codants) seulement.

Cette abondance de séquences non codantes est courante dans les génomes eucaryotes, d'autant plus que ceux-ci sont de grande taille.

2. De nombreuses séquences répétées

On réalise des expériences de dénaturation / renaturation de l'ADN (on utilise pour cela des fragments de 1000 bp, pour pouvoir comparer les résultats obtenus). Ces expériences sont basées sur la nature bicaténaire de l'ADN : La dénaturation est réalisée à 70°C (séparation des brins sans hydrolyse), puis on reforme l'ADN double brin à 20-25°C ; cette renaturation est suivie par l'absorbance aux UV, supérieure pour l'ADN simple brin (dénaturé).

Ces expériences permettent d'étudier deux aspects du contenu des génomes :

- La **taille** des génomes, proportionnelle au $Cot_{1/2}$
- La présence de **séquences en plusieurs exemplaires**, qui se réassocient plus rapidement

On confirme ainsi que les génomes eucaryotes ont une taille plus importante que les procaryotes, eux-mêmes plus importants que les virus.

On note toutefois la présence chez les **eucaryotes**, de **trois types de séquences** :

- Des séquences **très répétées** $\Rightarrow$ **10% min**
- Des séquences **moyennement répétées** $\Rightarrow$ **20-80%**
- Des séquences **uniques**, en un seul exemplaire

Alors que le génome des procaryotes ne présente que peu de séquences répétées, ces dernières sont donc très abondantes chez les eucaryotes.

3. La diversité de l'ADN non codant

L'ADN non codant est très diversifié, et peut présenter des rôles fonctionnels, structuraux ou évolutifs.

a. ADN hautement répété : l'ADN satellite

L'ADN **satellite** correspond à des fragments chromosomiques formés par de nombreuses répétitions d'une même séquence.

On y trouve par exemple l'**ADN télomérique** (qui permet en particulier la protection des extrémités des chromosomes), ainsi que l'**ADN centromérique** (qui permet en particulier la fixation des protéines des kinétochores pendant les divisions cellulaires).

Les autres séquences satellites peuvent être formées de répétitions de séquences plus ou moins longues, de 2 à 4 pb pour les microsatellites, jusqu'à 100 pb pour les minisatellites (ces derniers présentent une forte variabilité individuelle, et sont ainsi utilisés pour la détermination des empreintes génétiques).

b. ADN moyennement répété

On y trouve des séquences non codantes, correspondant à des **éléments génétiques mobiles**. Ces éléments peuvent représenter une part importante du génome (plus de 30% dans le génome humain, dont environ 10% pour le seul transposon *Alu*). Ces éléments mobiles peuvent potentiellement générer de la diversité génomique.

On trouve aussi des séquences codantes moyennement répétées : Il s'agit de **répétitions en tandem**, en particulier pour les gènes codant les ARN à rôle fonctionnel : ARNr, ARNt et petits ARN. On y trouve aussi les gènes codant les histones, là encore répétés en tandem.

c. Séquences peu ou pas répétées

On trouve dans le génome des pseudogènes, qui correspondent à des duplications de gènes ayant conduit à une copie non fonctionnelle.

Longtemps considéré comme « ADN poubelle », l'ADN non codant est de plus en plus considéré comme pouvant présenter un rôle important, en particulier au niveau du contrôle de l'expression génétique.

4. L'existence de gènes morcelés

De nombreux gènes eucaryotes sont morcelés : l'ARN pré-messager est constitué d'introns et d'exons. Les **introns** sont des séquences éliminées (excisées) lors de la maturation des ARNm, et les **exons** sont épissés (associés bout à bout).

Cet épissage se réalise grâce à des **sites spécifiques** situés aux bordures 5' et 3' de chaque exon : on parle d'épissage constitutif. Seuls les exons sont donc des séquences codantes, les introns faisant partie de l'ADN non codant.

En moyenne, les introns occupent chez les eucaryotes supérieurs 10 fois plus d'ADN que les exons. Chez les tétrapodes, la plupart des gènes sont des gènes avec une **structure mosaïque**.

Ils permettent de même un brassage des exons lors de l'expression de l'information génétique : on parle alors d'**épissage alternatif**. Il s'agit du fait qu'il est possible de ne réassocier que certains des exons d'un ARNprém ; en épissant différents exons, on peut ainsi aboutir à différents ARNm et donc à différentes protéines.

Chez l'Homme, plus de 35% des gènes subissent un épissage alternatif : formation d'environ 100 000 protéines, à partir de moins de 23 000 gènes !

Exemples : localisation de la protéine (IG membranaire ou sécrétée), substrat/ligand reconnu, ou protéines distinctes (calcitonine dans la thyroïde ; neuromédiateur CGRP dans le cerveau)...

IV. Les génomes viraux

Les virus ne sont pas des cellules, mais des **parasites obligatoires des cellules**. Ils sont capables de détourner la machinerie cellulaire afin de se multiplier.

Un virus est constitué d'une information génétique. Lors de la phase extracellulaire de son cycle, cette information génétique est associée à des protéines (et parfois une enveloppe lipidique), formant ainsi le **virion**, ou **particule virale**.

A. Un virus est un agent infectieux des cellules

1. Phase extracellulaire : le virion

La **particule virale** est constituée :

- Du matériel génétique
- *Eventuellement de protéines ou enzymes associées à ce matériel génétique*
- D'une capsid = coque protéique, formée de sous-unités répétées (les capsomères)
- *Eventuellement d'une enveloppe = bicouche lipidique et protéines, entourant la capsid*

2. Phase intracellulaire : le génome viral

La particule virale permet l'entrée du matériel génétique dans une cellule hôte, au sein de laquelle cette information génétique s'exprime, ce qui permet la production de nombreuses particules virales.

Le matériel génétique peut être présent sous différentes formes :

- ADN bicaténaire ; exemples : bactériophages T pairs, **phage λ**, adénovirus (angine, diarrhée...), herpès, variole, etc.
- ADN monocaténaire ;
- ARN bicaténaire ; exemple : rotavirus (gastro-entérite)
- ARN monocaténaire :
 - ARN (+) : l'ARN est équivalent à un ARNm, dont traductible ; exemples : **VMT**, **coronavirus**, poliomyélite, rubéole...
 - ARN (-) : l'ARN ne peut être traduit en protéine. On y trouve en particulier les rétrovirus, qui grâce à une reverse transcriptase synthétisent un ADNc qui s'intègre dans le génome de la cellule hôte ; exemples : grippe, VIH

Les génomes des virus présentent souvent des particularités inhabituelles au niveau de leur organisation, correspondant à une très forte compaction de l'information :

- Présence de cadres de lecture se chevauchant : une même séquence peut être lue de manière « décalée » pour former des protéines différentes.
- Synthèse de précurseurs protéiques de grande taille, qui sont ensuite clivés pour former plusieurs protéines différentes.

B. Un virus des bactéries : le phage λ

1. Le virion

C'est un virus de grande taille 150 à 200 nm environ qui présente une structure complexe. La tête fait 50 nm. C'est un assemblage protéique qui contient un ADN de 48,5 kb codant une soixantaine d'enzymes.

Le cylindre creux est constitué d'un grand nombre de protéines identiques disposées en anneaux superposés.

- Tête + queue = capsid ;
- Capsid + ADN = nucléocapsid.

2. Pénétration dans la bactérie

Il se fixe à une protéine spécifique de la surface de la bactérie hôte. Chez *E.coli*, cette protéine est une perméase permettant l'absorption du mannose par la cellule. Suite à cette fixation, la capsid reste en surface et seul l'ADN viral est injecté dans la cellule à travers la queue cylindrique creuse du bactériophage. Dans la cellule l'ADN se circularise.

3. Cycle du virus

Les virus qui attaquent les cellules bactériennes, les **bactériophages**, se multiplient dans la cellule bactérienne en commandant à la machinerie biochimique de la cellule de synthétiser de nouvelles copies du génome viral et de synthétiser les protéines de l'enveloppe virale. Le système génétique du phage λ est devenu l'un des systèmes les mieux compris.

On a mis cependant en évidence l'existence d'une latence virale lorsqu'on observa qu'à l'exposition à la lumière violette des bactéries apparemment non infectées étaient capables de produire des bactériophages. Des expériences ultérieures montrèrent que ces bactéries **lysogènes** portaient dans leur chromosome un ADN viral complet mais **dormant**.

Il existe ainsi deux cycles différents pour le phage λ :

- Dans le cas d'une **infection lytique** (phase **lytique**), les gènes phagiques de la reproduction sont actifs. Le phage produit les enzymes dont il a besoin pour orienter la cellule vers la production des éléments nécessaires à sa maturation et à la lyse de la cellule. → **Cycle lytique**
- Plus rarement, dans l'état **lysogène**, le génome du prophage est inactif (= réprimé). Le phage s'intègre dans le chromosome bactérien sous forme d'un prophage inerte (ce qui rend la bactérie **lysogène**). L'ADN lambda intégré est appelé **prophage** et il reste à l'état quiescent dans le chromosome bactérien. La bactérie infectée se multiplie normalement et transmet l'ADN lambda à tous ses descendants → de cette façon, l'ADN viral se multiplie de nombreuses fois sans détruire sa cellule hôte. → **Cycle lysogène**.

4. Le génome du virus

Parmi les gènes présents dans le génome du phage λ , on trouve en particulier :

- Des gènes codant les protéines de la **capside**.
- Une **réplicase** : ADN polymérase permettant la réplication de l'ADN viral en grandes quantités.

C. Un virus des plantes : le VMT

1. Le virion

Le **Virus de la Mosaïque du Tabac** (VMT) infecte différents Angiospermes (dont le Tabac *Nicotiana tabacum*, une Solanacée) et y déclenche une maladie caractérisée par l'apparition de plages colorées sur les feuilles, d'où le nom de mosaïque. Ces plages colorées correspondent à des zones de nécrose cellulaire.

Le virus apparaît comme un bâtonnet cylindrique de 300nm de long pour 17 nm de diamètre.

Ce bâtonnet est constitué d'une seule protéine (masse moléculaire 17,5kD, 158aa) présente en de nombreux exemplaires (2130)

L'information génétique est portée par une seule molécule linéaire d'ARN monobrin de 6930 nucléotides portant un génome limité à 4 gènes (ARN(+))

L'ensemble est organisé en structure hélicoïdale dans laquelle l'ARN est entouré en une hélice de 130 spires sur laquelle sont arrimés les capsomères formant la capsid du virus (chaque capsomère recouvre 3 nucléotides).

2. La pénétration dans la cellule végétale

Ce virus ne possède pas d'enzymes permettant la destruction des parois cellulaires. L'infection des cellules hôte se fait par des blessures, des cassures des feuilles. Une feuille intacte plongée dans une suspension de virus ne s'infecte pas. La contamination peut également être réalisée par des pucerons qui jouent alors le rôle de vecteurs.

3. Cycle du virus

Le cycle du VMT est **rapide** : 1 à 3 jours et se déroule dans les tissus métaboliquement actifs : le mésophylle principalement. Ce cycle aboutit à la mort des cellules hôte : c'est donc un cycle **lytique**.

La pénétration dans la cellule hôte est suivie de la **décapsidation** : la dissociation de la capsid libère l'ARN génomique. Celui-ci est qualifié **d'ARN(+)** car il joue le rôle **d'ARNm**.

Dans la cellule hôte, il est **directement traduit** en protéines virales à l'aide des ribosomes, des ARNt et des enzymes des cellules de feuille de tabac.

Cette traduction aboutie à la synthèse des **capsomères**, de **protéines inhibitrices des transcriptions** de la cellule et d'une enzyme virale absente de la capsid : la **réplicase**.

La réplicase est une ARN polymérase : l'ARN(+) joue le rôle de matrice et permet la synthèse d'un brin complémentaire : **ARN(-)**. Ces deux ARN sont unis en une double hélice : **forme répllicative**.

La réplicase catalyse une deuxième réplication utilisant maintenant l'ARN(-) → les brins formés sont des ARN(+). L'encapsidation de ces ARN forme de nouvelles particules virales.

Les sous-unités du virus peuvent s'assembler spontanément en tube à essai ; Le VMT fut le premier virus dont on ait démontré la capacité d'autoassemblage à partir de ses composants.

Lorsque le virus a pénétré dans une cellule végétale et qu'il commence à se reproduire, les particules virales se répandent dans l'ensemble de la plante en passant par les plasmodesmes, les canaux cytoplasmiques qui traversent les parois.

4. Le génome du virus

Le génome du VMT est extrêmement réduit.

Un premier gène permet la synthèse des deux protéines formant les sous-unités de la Réplicase : Les deux protéines sont traduites à partir du même codon initiateur, mais la plus grande est obtenue en « sautant » le codon STOP marquant la fin de la première protéine. Cette réplicase est une ARN polymérase.

Les deux autres gènes codent :

- La protéine de capsid (CP)
- Une protéine de mouvement (MP), qui permet le déplacement des particules dans la plante, en particulier au niveau des plasmodesmes.

D. Un virus humain : le SARS-Cov-2

1. Le virion

Les coronavirus sont des virus enveloppés : Ils possèdent une enveloppe formée d'une bicouche lipidique (héritée de la cellule infectée) et des protéines.

La nucléocapside comporte un ARN associé directement à des protéines.

Les coronavirus possèdent plusieurs membres responsables d'infections respiratoires chez l'Homme. Il s'agit pour certains de simples rhumes hivernaux, alors que des coronavirus apparus récemment (au XXIème siècle), de la famille des sarbecovirus, sont responsables d'épidémies plus ou moins étendues :

- SRAS (syndrome respiratoire aigu) en 2002-2003 (virus SARS-Cov-1)
- Syndrome respiratoire du moyen-orient en 2012 (virus MERS-Cov)
- COVID depuis 2020 (virus SARS-Cov-2)

2. La pénétration dans la cellule humaine

Le virus se fixe sur les cellules humaines grâce, en particulier, à une interaction entre la protéine **Spike** portée par son enveloppe et la protéine **ACE2** présente sur la face extracellulaire de certaines cellules (poumons, artères, foie, cœur, reins...).

Cette fixation permet la fusion de l'enveloppe virale avec la membrane plasmique, ce qui permet l'entrée de la nucléocapside dans le cytoplasme.

3. Cycle du virus

Une fois dans le cytoplasme, la nucléocapside se dissocie entre protéine N et ARN.

La **protéine N** joue un rôle inhibiteur sur la synthèse d'interféron par la cellule ; or l'interféron est une protéine sécrétée par les cellules infectées par des virus et permettant en temps normal l'activation d'une réponse immunitaire contre cette infection.

L'ARN est directement traductible :

- Les deux ORF 1a et 1b, traduites différemment par décalage du cadre de lecture, permettent la synthèse (après clivage protéique) de protéiques à rôle fonctionnel, dont la réplicase (qui est une ARN polymérase).
- Les autres gènes permettent la synthèse de protéines accessoires, et des protéines de la particule virale :
 - **Protéine S** = spike, formant des spicules sur l'enveloppe
 - Protéine N, formant la nucléocapside avec les ARN viraux néosynthétisés
 - Protéines E et M, dans l'enveloppe virale

Une fois assemblées, les nucléocapsides **bourgeonnent** en emportant un fragment de membrane comprenant les protéines S, M et E, ce qui permet donc la formation de nouveaux virions. Les protéines M et E semblent jouer un rôle important dans l'association entre nucléocapside et membrane, lors de ce processus.

La présence de protéine S à la surface de la cellule infectée semble aussi induire la fusion entre des cellules infectées, qui forment alors un syncytium (cellule plurinucléée), qui serait plus efficace pour la production de virions et plus difficile à éliminer par le système immunitaire.

4. Le génome du virus

Le génome des coronavirus présente donc deux ORF chevauchantes par décalage du cadre de lecture.

Comme chez de nombreux virus, ces ORF codent des précurseurs protéiques de grande taille, clivés secondairement afin de former les différentes protéines virales.

Le génome possède au final 4 gènes de structure (formant la nucléocapside et l'enveloppe), 2 gènes à rôle fonctionnels (conduisant à la production de 11 et 16 protéines, dont l'ARN polymérase) et 9 gènes accessoires.