

LA CONSTITUTION GENOMIQUE : STABILITE, VARIABILITE *

J.Cl. Manuguerra ¹

RESUME : Dans leur adaptation à leur environnement et à sa conquête, les agents pathogènes conventionnels, dont le patrimoine génétique est consigné dans une ou plusieurs molécules d'acides nucléiques, utilisent trois grands mécanismes de variations. Un mécanisme est continu et progressif, c'est la mutation ponctuelle. Elle correspond au changement d'un nucléotide par un autre et est liée à l'infidélité des éléments qui interviennent dans la multiplication du matériel génétique. Les mutations peuvent s'accumuler dans le temps et constituent ainsi une base de l'évolution des constitutants viraux et bactériens. En marge des mutations, deux événements peuvent également se produire au moment de la duplication du matériel génétique : c'est tout d'abord la soustraction d'une partie de l'information (appelée délétion) ou l'addition d'éléments génétiques (baptisée insertion). Le deuxième grand mécanisme de variation est le réassortiment génétique chez les virus, que l'on peut comparer à l'acquisition ou à l'échange de plasmides chez les bactéries. Ce phénomène est discret et peut conduire à l'acquisition brutale de nouvelles propriétés phénotypiques (changement radical de profil antigénique, de virulence ou acquisition d'une résistance aux antibiotiques). Enfin, la recombinaison, qui existe chez les bactéries et chez de nombreux virus, représente un échange de matériel génétique qui est un puissant moteur d'évolution génomique. L'accès à la connaissance du patrimoine génétique des agents pathogènes conventionnels et le suivi de leurs variations génomiques permet :

1. de détecter et de typer des agents pathogènes, généralement en ciblant des structures stables ;
2. d'étudier l'information elle-même au niveau génétique et au niveau protéique par déduction.

Dans ce dernier cas, les relations de proximité ou les différences sont plutôt étudiées au niveau de structures variables, voire hypervariables. Les analyses portant sur le génome doivent s'interpréter en fonction de l'agent pathogène lui-même et de la stratégie technique choisie, notamment.

SUMMARY : In their quest for adaptation to their environment and their conquest of new « territories », conventional pathogens, whose heredity lies within one or many molecules of nucleic acid, undergo three main mechanisms of variations. One of these, called mutation, is continuous and progressive. It corresponds to the change of a nucleotide by an other and is associated with the infidelity of the element involved in the duplication of the genetic material. Mutations may cumulate in the course of time and are therefore the basis of the evolution of one or many viral or bacterial component(s). Besides mutations, two events may also occur during the phase of duplication of the genes: one is the deletion of nucleotides, the other is an insertion of nucleotide. The second mechanism of variation is the genetic reassortment, in the case of viruses, which can be compared with the acquisition or the exchange of plasmids in the case of bacteria. This phenomenon is discrete and may lead to the brutal acquisition of new phenotypic properties (radical change in virulence or in an antigenic profile or acquisition of resistance to antibiotics). Lastly, genetic recombination, the third mechanism of genetic variations, is a powerful drive for the evolution of bacteria and some viruses. The access to the genomes of conventional pathogens and to their variations allows :

1. to detect and type these agents generally by targeting stable or poorly variable structures;
2. to analyse the genetic information itself and the deduced protein data by targeting variable, or even highly variable, genes.

Whatever genetic analysis is undertaken, it must be interpreted considering, among other factors, the nature of the pathogen itself and the technical strategy used to get results.



* Texte de l'exposé présenté lors de la journée AEEMA, 17 mai 2001

¹ Unité de Génétique Moléculaire des Virus Respiratoires, Institut Pasteur, F-75724 Paris cedex 15, France

I - NATURE DU SUPPORT DE L'INFORMATION

Les agents pathogènes « conventionnels » disposent tous d'un patrimoine héréditaire : qu'il s'agisse de parasites unicellulaires, de champignons microscopiques, de levures, de bactéries ou de virus. Pour la simplicité de cette communication, les exemples seront centrés sur les virus et quelquefois sur les bactéries, pour lesquelles cependant des exemples plus détaillés seront évoqués ci-après dans la communication de Henri-Jean Boulouis (page 21). Le patrimoine héréditaire est consigné dans des molécules d'acides nucléiques. Comme pour celui des organismes eucaryotes (c'est à dire pourvus d'un véritable noyau cellulaire) plus complexes, champignons et levures, le patrimoine génétique des bactéries (procaryotes, c'est à dire sans noyau véritable) est contenu dans de l'ADN double brin appelé chromosome bactérien. En plus de cet ADN, qui est le support du génome des bactéries, ces dernières sont pourvues d'autres acides nucléiques : les ARN. Certains d'entre eux sont seulement « temporaires » et constituent des intermédiaires dans le flux de l'information du génome vers les protéines, d'autres en revanche sont « constitutifs » et contribuent à l'édification d'organelles intracellulaires. Il est également important de noter qu'en plus de l'ADN chromosomique bactérien, il existe chez les bactéries un autre type de molécules d'ADN double brin : l'ADN plasmidique. Ces structures portent un certain nombre de caractères et sont notamment impliqués dans des mécanismes de résistance aux antibiotiques. Les bactéries contiennent tous les éléments nécessaires au métabolisme cellulaire, dont il est possible de dresser un profil, utilisable pour leur caractérisation. Les virus, quant à eux, ne contiennent aucun de ces éléments ; ce sont des parasites obligatoires qui « profitent » du métabolisme de la cellule qu'ils infectent. Les virus ne contiennent qu'un seul type d'acide nucléique : soit de l'ADN, soit de l'ARN. Dans un cas comme dans l'autre, le génome viral peut être simple brin ou double brin. Pour les virus à ARN, ou ribovirus, l'ARN est soit positif ou infectieux (il est messager, c'est à dire qu'il peut être directement utilisé par la cellule hôte pour conduire la synthèse de protéines), soit négatif ou non infectieux (il doit être copié en un brin complémentaire au moins partiel avant qu'il puisse diriger la synthèse de protéines).

Malgré leurs différences fondamentales, les bactéries et les virus ont ceci en commun que leur génome ou autres acides nucléiques sont protégés de la dégradation dans le milieu extérieur par des structures protéiques ou glycolipidoprotéiques : paroi, membrane, capsid, enveloppe. En plus de leur rôle protecteur, ces structures externes sont douées de fonctions, comme celle d'attachement par exemple qui est nécessaire à la fixation à un récepteur cellulaire et qui, pour les parasites intracellulaires, est indispensable à l'entrée de la bactérie ou du virus dans la cellule cible.

Pour les bactéries, l'existence d'une paroi de peptidoglycane ou son absence fournit un élément de caractérisation par coloration chimique. C'est le cas de la coloration de Gram, qui permet de classer les bactéries en deux catégories : les GRAM positives et les GRAM négatives. Cette classification s'ajoute aux critères d'identification des procaryotes qui incluent : la forme, la façon dont les bactéries se groupent et leur motilité.

Des structures bactériennes externes assurent l'attachement de virus qui infectent les bactéries. Ces virus ou bactériophages sont utilisés pour caractériser les bactéries selon qu'ils les lysent ou non, c'est la lysotypie. Enfin, les parois et membranes bactériennes déterminent des antigènes qui suscitent une réponse en anticorps et une réponse immunitaire spécifique à médiation cellulaire. Les réactions antigènes et anticorps spécifiques sont la base des méthodes de caractérisations immunologiques : c'est la sérotypie.

Les virus, quant à eux, sont soit nus, c'est-à-dire pourvus d'une seule caspide de symétrie hélicoïdale ou cubique, soit enveloppés. Dans ce cas, leur(s) nucléocapside(s) est(sont) contenue(s) dans une particule virale délimitée par une bicouche lipidique, empruntée à la cellule hôte, sous-tendue ou non par une couche de protéine matricielle. Cette enveloppe est couverte ou hérissée de glycoprotéines, impliquées pour certaines d'entre elles dans l'entrée du virus (attachement puis fusion de l'enveloppe virale avec la membrane cellulaire), qui portent également des motifs antigéniques. Dans le cas des virus aussi, les réactions antigènes et anticorps spécifiques sont la base des méthodes de caractérisation immunologique de sérotypage.

A la complexité de structure des agents pathogènes conventionnels correspond une taille de génome plus ou moins grande (tableau I). Si la taille du génome des champignons microscopiques et levures est de l'ordre de 10 600 000 paires de bases, celui des bactéries comme *Echerichia coli* a une taille de l'ordre de 4 700 000 paires de bases. Les virus sont beaucoup plus petits. Leur génome est généralement très compact et des stratégies de codage permettent au virus de faire des « économies » de nucléotides. Le génome des gros virus à ADN, comme ceux de la famille des *Poxviridae* (virus de la variole humaine, virus de la vaccine par exemple) est de l'ordre de 200 000 paires de bases, tandis que de tout petits virus comme les membres de la famille des *Picornaviridae* (virus de la poliomyélite aiguë paralytique chez l'Homme, virus de la fièvre aphteuse chez les animaux à pied fourchus, virus de la maladie vésiculeuse des suidés) ont un génome de l'ordre de 7 500 bases seulement (tableau I). Compte tenu de la taille respective de ces 'microorganismes', l'analyse du génome fera appel à des techniques soit différentes (technique du champ pulsé pour les

bactéries, non utilisée en virologie) ou identiques mais appliquées à des structures ou gènes différents (amplification génique en chaîne ou PCR pour la mise

en évidence d'un gène particulier ou d'une mutation particulière au sein d'un gène particulier).

TABLEAU I
Diversité du génome parmi les « micro-organismes »

Type de 'micro-organisme'	Exemples	Support de l'information génétique	Taille du génome et « produits »	Source
Champignons microscopiques/ levures	<i>Saccharomyces cerevisiae</i>	ADN double brin linéaire en 16 chromosomes	16 chromosomes 12058095 pb	Taille du génome calculée d'après http://genome-www.stanford.edu/Saccharomyces/MAP/GENOMICVIEW/GenomicView.html
Bactéries				
	<i>Escherichia coli</i>	<i>Escherichia coli</i> K12 ADN double brin circulaire	Chromosome bactérien 4639221 pb Protein coding genes: 4289 Structural RNAs: 115	GenBank: numéro d'accèsion U00096
	<i>Mycobacterium tuberculosis</i>	<i>Mycobacterium tuberculosis H37Rv</i> ADN double brin circulaire	Chromosome bactérien 4411529 bp Protein coding genes: 3927 Structural RNAs: 51	GenBank: numéro d'accèsion AL123456
Virus				
<i>Poxviridae</i>				
	Virus de la vaccine	Souche Copenhagen ADN double brin	191636 b	GenBank: numéro d'accèsion M35027
<i>Picornaviridae</i>				
	Virus de la poliomyélite aiguë paralytique humaine	Poliovirus de type 1 ARN positif simple brin non segmenté linéaire	7 441 b Une polyprotéine clivée suite à la traduction	GenBank: numéro d'accèsion V01150; J02282; J02285; J02286; V01133
	Virus de la fièvre aphteuse	Type O ARN positif simple brin non segmenté linéaire	7 712 b Une polyprotéine clivée suite à la traduction	GenBank: numéro d'accèsion M10975

II - MOYENS D'INVESTIGATION DU MATERIEL GENETIQUE, CHOIX DE LA CIBLE A ETUDIER

Actuellement, l'épidémiologie moléculaire repose pour une part importante sur l'analyse de séquences soit d'acides nucléiques, soit de séquences d'acides aminés qui en sont déduites, déterminées à partir du matériel génétique issu des bactéries ou des virus. Les séquences analysées sont soit une succession de lettres prises dans un alphabet de quatre lettres (A/T ou U/G/C) pour les séquences nucléotidiques ou dans un alphabet de vingt et une lettres, pour les séquences

protéiques. La comparaison de séquences porte le plus souvent sur un gène choisi pour des caractères particuliers, qui seront discutés ci-après. La comparaison se fait en plusieurs étapes. Tout d'abord, il y a une étape de synthèse d'acide nucléique, le plus souvent par amplification génique ou PCR, précédée d'une étape de transcription inverse dans le cas des virus à ARN. Cette étape d'amplification peut servir à la détection de l'organisme recherché et à son identification, au moins partielle. A partir du matériel

ainsi généré, une réaction de séquençage, proprement dit, est ensuite réalisée par la méthode des di-déoxy nucléotides. Une fois la séquence obtenue, elle est comparée à l'aide de logiciels informatiques conçus autour d'algorithmes. Le fruit de l'alignement des séquences entre elles et de leurs scores de ressemblance peut être représenté par un arbre phylogénique, sorte d'arbre généalogique. Ce type de représentation permet de visualiser la proximité génétique de souches bactériennes ou virales.

La comparaison de séquences et la construction d'arbres phylogéniques reposent sur des méthodes dont les principales, certaines exhaustives d'autres aléatoires, sont les suivantes :

- Méthode de maximum de parcimonie et parcimonie pondérée (*méthode : état de caractères, recherche exhaustive*)
- Méthode de maximum de vraisemblance (*méthode : état de caractères, recherche exhaustive*)
- Méthodes UPGMA / WPGMA (*méthode : matrices de distance, groupements par étapes*) [U(W)PGMA = Unweighted (weighted) pair group method with arithmetic means]
- Méthode de Fitch et Margoliash (*méthode : matrices de distance, recherche exhaustive*)
- Méthode du plus proche voisin (*méthode : matrices de distance, groupements par étapes*).

L'interprétation de l'arbre phylogénique peut être trompeuse si l'on tient compte d'une simple impression visuelle non corrigée par la connaissance du fait de base suivant : seules les distances horizontales comptent ; l'ordre vertical et le choix de l'enracinement de l'arbre sont arbitraires. Un autre paramètre à observer, et qui manque souvent, est le nombre de réitérations dit de '*bootstraps*' qui reflètent la robustesse de chaque nœud de l'arbre et globalement de l'arbre lui-même. Plus cette valeur est élevée plus l'arbre est robuste et moins il risque d'être le fruit du hasard.

Le choix du gène séquencé, voire même de la région du gène séquencé, peut donner des résultats très différents (voir ci-après). Leur interprétation doit donc en tenir compte. Le choix du gène pour le séquençage ou tout simplement pour la détection et l'identification dépend de la question posée et du but recherché. Ainsi, le choix s'orientera plutôt vers des séquences codantes (qui dirigent la synthèse de protéines) ou plutôt vers des séquences non codantes (séquences de régulation). On s'orientera soit vers des séquences conservées soit vers des séquences variables voire hypervariables. On peut aussi choisir des séquences répétées dont on étudie le schéma de la répétition. Pour les séquences codantes, on peut choisir des gènes qui dirigent la synthèse de protéines de structure ou bien de protéines non structurales.

Pour la caractérisation d'un virus, par exemple du virus de l'immunodéficience humaine de type 1 (VIH-1), on prend comme cible de l'analyse moléculaire des gènes impliqués dans la synthèse de protéines de structure et plus particulièrement celui qui dirigent la synthèse des protéines d'enveloppe : les glycoprotéines GP160. Ces cibles permettent de déterminer qu'il s'agit de VIH, qu'il est de type 1 et on peut même aller jusqu'au groupe ou sous-type. Pour l'amplification de la cible de l'étude par RT-PCR, on évite soigneusement de placer des amorces spécifiques nécessaires au démarrage des réactions d'amplification dans la boucle V3 hypervariable de la GP 160. En effet, des variations de séquences au niveau ciblé par les amorces entraîne un défaut de reconnaissance qui aboutit à l'absence de signal. Ceci peut conduire à un résultat de détection/identification qui est un faux négatif.

Le choix de la cible pour la détection/identification par les techniques de génétique moléculaire est associé à des niveaux d'informations différents mais également à un spectre de détection variable. Ainsi, pour les virus grippaux par exemple, cibler un gène dirigeant la synthèse d'une protéine interne comme celui de la nucléoprotéine permet de détecter un grand nombre de virus et donne comme information qu'un virus grippal est présent et qu'il appartient à un type considéré. Dans ce cas, il est nécessaire d'introduire dans la réaction des réactifs qui permettent de reconnaître soit des virus de type A, soit des virus de type B ou encore des virus de type C. Il ne sera pas possible d'aller plus loin dans l'identification du virus. Pour ce faire, il faut cibler le gène de l'hémagglutinine et de la neuraminidase grippales pour déterminer le sous-type complet auquel le virus appartient. On peut éventuellement aller jusqu'à un génotypage. Si la détection de la cible va au delà de l'identification par RT-PCR et est prolongée par une étude de la séquence, le choix des gènes est primordial. Ainsi, comme cela a été signalé précédemment, une étude phylogénique peut apporter des résultats différents en fonction du gène choisi ou de la région du gène choisie. Prenons le cas de la grippe chez le porc. L'analyse consécutive à une détection/identification par RT-PCR ou de RT-PCR suivi de séquençage de la région amplifiée, qui porte seulement sur l'hémagglutinine (HA) qu'un virus porcin de grippe A(H1N1) européen, conduit à identifier ce virus comme virus aviaire. En effet, la glycoprotéine de surface H1 de ces virus a été empruntée à des virus d'oiseaux. En revanche, si les mêmes analyses portent sur les séquences nucléiques ou protéiques associées à la nucléoprotéine de ce même virus porcin, le résultat sera qu'il s'agit d'un virus A(H1N1) porcin classique. La confrontation des deux résultats complémentaires est nécessaire pour véritablement caractériser le virus grippal en cause. C'est le fruit d'un réassortiment génétique (mécanisme détaillé ci-après) relativement récent. Une analyse qui porte seulement sur la molécule H1 suffit à sous-typé le virus et est pertinente en terme vaccinal mais son résultat ne reflète qu'une partie du génome et de

l'origine de ces virus et est insuffisant en terme d'épidémiologie (moléculaire).

Le choix de la cible de l'étude et la technique mise en œuvre doivent toujours être adaptés à la question dans un contexte donné. Par exemple, la séquence de l'extrémité 5' non codante du génome des virus de la famille des *Picornaviridae*, est conservée et peut servir de cible à la détection par réaction de RT-PCR. Dans ce cas, la réponse est : absence ou présence d'un membre de cette famille dans le prélèvement. Si l'on évoque ce résultat, dans un contexte de paralysie aiguë virale chez l'Homme dans des régions ou des communautés où le virus de la poliomyélite n'a pas encore été sûrement éradiqué, il est possible de conclure avec un degré élevé de confiance qu'il s'agit d'une infection par un poliovirus. En revanche, il n'est pas possible de dire s'il s'agit d'une infection par un poliovirus sauvage ou, si en cas d'usage d'un vaccin vivant oral, s'il s'agit d'une maladie consécutive à la

réversion vers la virulence d'un des trois virus vaccinaux.

Parallèlement, dans un contexte d'épizootie de fièvre aphteuse, si l'on considère ce même résultat pour un mouton porteur de lésions aphteuses, il est possible de considérer le diagnostic de fièvre aphteuse pour acquis avec un très bon niveau de confiance. Si on contraire, un mouton présente les mêmes lésions avec un résultat du test de RT-PCR sur la région 5' non codante, hors d'un contexte épizootique, le diagnostic de fièvre aphteuse ne reposera pas sur le même niveau de confiance. Toujours sur ce même exemple, dans le cas de lésions vésiculeuses de type aphteux chez le porc, pendant ou en dehors d'un contexte épizootique, un résultat de RT-PCR positive dans la région 5' non codante du génome viral ne permet pas de conclure qu'il s'agit d'une infection par un virus aphteux ou par un virus de la maladie vésiculeuse des suidés qui est proche voire identique au virus Coxsackie B5, un entérovirus qui infecte l'Homme.

III - MECANISMES DE VARIATION DU MATERIEL GENETIQUE

1. NIVEAUX D'INFORMATION ET DIVERSITE BIOLOGIQUE

La hiérarchisation des niveaux d'information reflète les niveaux de la classification taxinomique : ordre, famille, sous-famille, genre, espèce. Ces classifications reposent sur des caractères stables : données phénotypiques morpho-structurales et spectre d'hôtes quelquefois. L'espèce constitue un niveau taxinomique, homogène pour les caractères qui la détermine. Qu'il s'agisse d'une espèce bactérienne ou virale, au sein de celle-ci, il existe, entre ses membres, une diversité génétique qui peut être plus ou moins importante. On peut déterminer des types ou des sous-types, et au sein de ces derniers des variants. Pour les virus grippaux, la famille est *Orthomyxoviridae*, le genre est *Influenzavirus A* ou *B* ou *C* et pour le genre *Influenzavirus A*, il y a une espèce type A/Puerto Rico/8/34(H1N1). Au sein de ce genre de virus grippal aussi étiqueté type (type A), on distingue des sous-types en fonction des antigènes de surface H et N. Ainsi l'espèce prototype est de sous-type H1N1. Au sein de ce sous-type, on distingue plusieurs variants. Par exemple, le composant correspondant à ce sous-type dans le vaccin de l'année 1999-2000 était A/Beijing/262/95(HN1) tandis que celui de la saison suivante était le variant A/New Caledonia/20/99(H1N1). Ces deux variants avaient des propriétés antigéniques différentes. En réalité, même au sein d'un même variant, tout isolat est lui-même un mélange de populations virales qui détermine un ensemble baptisé quasi-espèce. Les caractères phénotypiques observés sont soit une résultante de la moyenne soit d'un caractère dominant. Cette moyenne

ou cette dominance sont liées à l'existence de génomes individuels dans le mélange qui constitue la quasi-espèce.

2. MUTATIONS PONCTUELLES

Les variations génétiques qui se répercutent ou non au niveau des caractères phénotypiques, évoluent selon des mécanismes soit discrets soit continus. Au registre des mécanismes d'évolution continue, il est possible de classer les mutations ponctuelles. Ces mutations sont dites ponctuelles car elles correspondent au changement d'un nucléotide dans la séquence d'un acide nucléique. Ce changement ou mutation est le plus souvent lié à une erreur de recopiage de l'acide nucléique. Les virus sont des organismes qui ont un taux de mutations qui peut être très fort. En effet, contrairement à des organismes plus complexes comme les bactéries et les cellules eucaryotes, les virus ne possèdent pas de « réparases », espèces d'enzymes qui repèrent les erreurs de recopiage après coup et les réparent. La plupart des virus se reproduisent grâce à une molécule de copiage qui leur est propre. Cette molécule qui est de type enzymatique et appelé polymérase ne possède pas de propriété d'édition, c'est à dire qu'elle ne repère pas les erreurs à mesure qu'elle assure la copie du brin parent en brin enfant. Pour ces raisons, les virus, surtout ceux à ARN, peuvent quelquefois évoluer très rapidement. Dans une étude portant sur 254 séquences du domaine HA1 de l'hémagglutinine H3 des virus humains isolés entre 1984 et 1996, il a été montré que le gène HA évoluait au taux de 5,7 substitutions nucléotidiques par an soit

$5,7 \times 10^{-3}$ substitutions par site et par an [1]. L'évolution des virus est le fruit de ces mutations qui surviennent au hasard et sont suivies ou non d'une sélection. Plus la pression de sélection est grande, plus l'évolution est rapide. Ainsi, généralement, les protéines virales externes sont soumises à une forte pression de sélection car elles sont des cibles de l'immunité, humorale le plus souvent. Ce sont donc les antigènes viraux « externes », qui sont généralement les plus variables. Au contraire, les antigènes internes généralement moins exposés évoluent moins rapidement.

Les mutations sont des changements « progressifs et continus ». Bien que ponctuelles, elles peuvent cependant entraîner des changements de phénotypes importants notamment au niveau de la virulence. Par exemple, pour les virus grippaux A(H5N1) d'origine aviaire mais isolés chez l'Homme lors de l'épisode dit de la grippe du poulet en 1997 à Hong Kong, le changement du résidu 627 sur la protéine PB2, un partenaire du complexe polymérase viral, est clairement impliqué dans l'augmentation de la virulence dans un modèle expérimental murin. Des mutations au niveau nucléotidique existent toujours. Ainsi, pour les virus à ARN, (comme les VIH-1 ou le virus de la fièvre aphteuse par exemple), pour lesquels les taux moyens de mutation sont de l'ordre de 10^{-4} à 10^{-5} (allant de 10^{-3} à 10^{-6}), en l'absence de capacité d'édition et de réparation, pour un génome de 7 500 à 10 000 nucléotides, il existe un mélange de génomes qui sont en compétition et éventuellement en évolution. La séquence observée, par séquençage de produits de PCR par exemple, est une séquence consensus, qui est une moyenne qui correspond à une séquence qui n'existe pas forcément. Dans ce mélange, il existe une séquence qui correspond au virus qui a la vitalité la meilleure et qui est appelée la séquence maîtresse. En cas de modification de milieu qui confère un avantage à un génome qui correspond à la séquence maîtresse, alors le virus évolue, pour le gène considéré, vers un état qui lui confèrera le plus de vitalité. Cette sélection aboutit à l'enrichissement progressif en séquence maîtresse et à la diminution de la séquence consensus qui peu à peu perd ce caractère pour être supplantée par la séquence maîtresse nouvelle qui s'impose comme nouveau consensus.

Il existe des situations où les mutations qui apparaissent sont délétères pour le virus qui les porte et ces virus sont alors contre-sélectionnés. On peut de plus, pour un même virus avoir une situation où l'apparition des mutations est bénéfique pour le virus, sur les antigènes externes par exemple (échappement à l'immunité), alors que d'autres mutations sont contre-productives pour les virus qui les porte, sur les antigènes internes par exemple (déterminant d'adaptation à l'hôte). Alors que pour les virus grippaux aviaires et humains, les taux de mutations et d'évolution des séquences nucléotidiques sont du même ordre, pour les premiers, l'évolution au niveau

des séquences protéiques de la nucléoprotéine est pratiquement nulle (stase évolutionnelle). Pour les seconds, au contraire l'évolution des séquences protéiques est rapide. Dans un cas, les virus aviaires sont adaptés depuis longtemps à leur hôte et un équilibre a été trouvé ; tout changement nuisant à cet équilibre conduit à une contre-sélection. Dans l'autre cas, les virus humains ont reçu leur nucléoprotéine grippale d'un autre hôte (l'oiseau *via* ou non le porc) depuis moins longtemps et un équilibre n'a pas encore été trouvé. Pour ces virus humains, à taux de mutation égale, l'évolution de l'hémagglutinine est encore plus grande que celle de la nucléoprotéine à cause de la pression de sélection encore plus forte sur cette protéine externe.

3. INSERTIONS ET DELETIONS DE MATERIEL GENETIQUE

A côté des mutations, il existe des « opérations » de soustraction de partie de séquence appelé délétions ou des « opérations » d'addition appelés insertions.

L'insertion d'acides aminés basiques au site de clivage de l'hémagglutinine du virus grippal pour les souches de virus aviaires de sous-type H5N2 est associée à une très forte augmentation de la virulence des virus. Dans le passé, elle a transformé de simples virus grippaux en virus de peste aviaire [2, 3].

4. LE REASSORTIMENT GENETIQUE

Un autre grand mécanisme de variation est le réassortiment. Il est possible seulement pour les organismes qui ont un génome porté par plusieurs molécules d'acides nucléiques physiquement indépendantes et correspond à l'échange d'une ou plusieurs de ces molécules, au moment d'une co-infection. Ceci aboutit au changement complet d'une ou plusieurs molécules protéiques dont la synthèse dépend du fragment nucléique échangé. Chez les virus grippaux, le réassortiment correspond au changement complet d'une molécule de surface telle que l'HA. La cassure antigénique *sensu stricto*, par remplacement de l'HA ou/ et de la neuraminidase (NA) par une HA ou/et une NA d'un type moléculaire différent, n'existe que pour le type A, au sein duquel elle aboutit à l'apparition de nouveaux sous-types, qui chez l'Homme peut conduire à une situation pandémique. Chez les bactéries, l'acquisition ou l'échange d'un ou plusieurs gènes, qui s'apparente au réassortiment, s'opère lorsqu'il y a échange de plasmides lors de la conjugaison. Ces échanges sont importants :

1. sur le plan antigénique pour les virus grippaux ;
2. sur le plan du phénotype résistant ou sensible aux antibiotiques des bactéries dans le cas de l'acquisition de plasmides.

L'analyse d'événements de réassortiment peut avoir une très grande importance en terme d'épidémiologie d'intervention. Ainsi, l'analyse phylogénétique des huit segments du génome des virus grippaux A(H5N1) isolés chez l'Homme à Hong Kong en 1997 [4], a révélé que tous étaient issus de virus établis chez les oiseaux et qu'aucun ne s'apparentait à des segments génomiques issus de virus inféodés à une espèce de mammifère (tableau II). De plus, les analyses phylogénétiques, complétées par des études sérotypiques ont indiqué qu'il existait deux lignées de virus humains A(H5N1) qui se superposaient à deux lignées de virus de poulets. Ces deux groupes viraux circulaient dans deux sous-ensembles géographiques de

Hong Kong et de son territoire. Au total, ces données signifiaient que les virus aviaires A(H5N1) étaient très probablement passés directement des poulets vivants à l'Homme très récemment, qu'ils étaient restés *purement aviaires* et étaient toujours peu adaptés à l'Homme. Seules des sources très contaminantes comme de nombreuses volailles vivantes fortement infectées étaient capables d'infecter des Hommes. Les hommes entre eux étaient incapables de se contaminer efficacement. Les virus A(H5N1) n'ont ainsi pas réussi leur implantation chez l'Homme et la suppression de la source contaminante (anadémie) a permis de faire disparaître le danger.

TABLEAU II
Analyse moléculaire des segments génomiques du virus A/Hong Kong/156/97(H5N1)

Segment génomique	Région amplifiée	Région Séquencée	Virus qui présente le plus fort degré d'identité	Sous-type	Identité (%)
PB2	8-2337	61-2310	A/Bugerigar/Hokkaido/1/77	H4N6	90,1
PB1	10-2340	32-2291 31-1754	A/Singapore/1/57	H2N2	90,6
PA	1-2200	& 1974-21798	A/Duck/Hokkaido/8/80	H3N8	91
HA	1-1773	21-1743	A/Turkey/England/91	H5N1	93,5
NP	7-1561	21-1542	A/Mallard/Astrakhan/244/82	H14N6	93,6
NA	1-1400	29-1391	A/Parot/Ulster/73	H7N1	91,1
M	775-1027	796-1001	A/Turkey/Minnesota/833/80	H1N1	98,5
NS	1-890	27-860	A/Duck/Hong Kong/717/79	H1N3	93,7

5. LA RECOMBINAISON GENETIQUE

Le dernier grand mécanisme de variation génétique est la recombinaison proprement dite qui implique l'échange de fragments d'ARN ou d'ADN entre deux brins nécessitant au moins un *crossing over*. Ce dernier mécanisme est fréquent et important pour des virus comme les VIH, coronavirus, poliovirus, herpesvirus et poxvirus. La co-infection avec deux souches virales différentes d'un même virus permet l'échange de matériel génétique. Ainsi, le long d'un génome, des morceaux de séquences nucléotidiques proviennent de l'un ou de l'autre virus. Les virus, qui sont dans ce cas, sont baptisés virus mosaïques.

Comme le montre la figure 1, des poliovirus de type 2 mosaïques ont été isolés : ils portent dans différentes régions de leur génome des séquences caractéristiques de la souche vaccinale Sabin et de virus sauvages polio de type 2 [5]. La connaissance de ce phénomène est très important dans le cadre d'une politique vaccinale fondée sur un recours exclusif au vaccin vivant atténué oral et dans la perspective de l'éradication de la poliomyélite virale paralytique aiguë et des poliovirus.

Dans le cas du virus de l'immunodéficience de type 1, la transmission d'un chirurgien à son patient a été démontré avec un très fort indice de confiance grâce à la comparaison des structures « mosaïques » des gènes de l'enveloppe (*env*) [6].

6. EVOLUTION DANS LE TEMPS ET DANS L'ESPACE

Le suivi des mutations dans le temps ou dans l'espace peut donner des informations précieuses sur l'origine d'un virus ou d'une bactérie. Pour le virus dont le taux de mutation se reflète au niveau du nombre de changements nucléotidiques, la connaissance des séquences nucléotidiques permet de disposer d'une « horloge moléculaire » qui rend possible une estimation de l'émergence d'un phénomène lié à la circulation d'une souche virale. Dans l'environnement, ceci peut être utile pour déceler une pollution d'origine humaine par la détection d'entérovirus, la détermination de leur 'séro type' et la « filiation » de souches virales par rapport à d'autres.

Les analyses combinées des mutations, insertions, délétions, réassortiments permettent aussi de suivre des variations dans le temps (évolution des hémagglutinines des virus grippaux H3 dans le temps depuis l'introduction du gène H3 à partir des canards quelque temps avant 1968) ou des variations dans l'espace. Ainsi, les alphavirus se répartissent en topotypes distincts qui recouvrent des aires géographiques définies. C'est le cas également des VIH-1, dont les groupes se répartissent différemment selon les continents. Le sous-type B est le plus largement distribué géographiquement. Il est

prépondérant ou exclusif sur l'ensemble du continent américain, en Europe et en Australie. Il est devancé en revanche en Inde par le sous-type C tandis qu'en Afrique la situation est plus complexe car il existe de multiples sous-types en co-circulation : notamment A, C D. Le sous-typage génétique des VIH de type 1 peut également se considérer selon les variations dans le

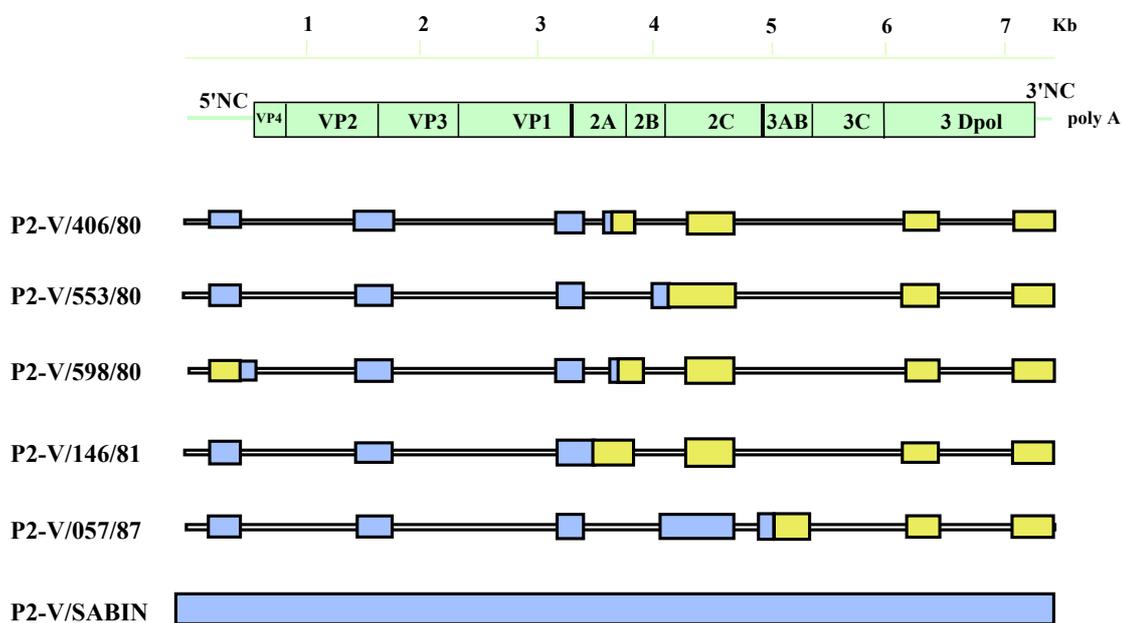
temps de la part relative de chaque sous-type. En dix ans, le sous-type C a gagné du terrain relativement aux autres puisqu'il représentait 8% des cas répertorié en 1989 et 53% en 1999. Cette évolution est liée à la dynamique épidémique de l'infection à VIH dans les pays asiatiques et africains.

FIGURE 1

Structure génomique des poliovirus recombinants sauvages et vaccinaux

La structure de chaque génome a été déduite d'analyses de profils obtenus après RT-PCR et digestion par enzymes de restriction ainsi de l'analyses de séquences partielles. Les portions de génomes séquencées sont indiquées par des rectangles. En foncé, les séquences de virus vaccinal de type 2 et en clair des séquences de poliovirus sauvages. Illustration empruntée à Sophie Guillot, Laboratoire d'épidémiologie moléculaire des entérovirus, Institut Pasteur, Paris.

[Source : « Natural genetic Exchanges between Vaccine and Wild Poliovirus Strains in Humans » 2000 ; Guillot *et al.*, Journal of Virology, Vol. 74, n° 18, 8434-8443.]



IV - CONCLUSION

Ainsi, le support de l'information génétique des bactéries et des virus permet, selon les techniques et les gènes ciblés, de détecter un agent pathogène, de le typer ou/et de le sous-typer. L'identification de type ou de sous-type repose surtout sur des éléments stables du génome. Une information plus précise repose plutôt sur des éléments variables du génome. Des exemples ci-avant ont illustré le fait que des variations ponctuelles permettent quelquefois de dater un événement ou de tracer une « filière » de transmission. D'autres exemples ont montré que des assortiments de séquences génétiques (recombinaisons ou réassortiments) peuvent indiquer l'origine d'un virus :

espèce hôte, « filière » voire individu. En tout état de cause, l'interprétation de résultats d'analyses génétiques doit tenir compte :

- de la nature de l'évènement épidémique ;
- de la nature de l'hôte ;
- de celle de l'agent pathogène ;
- et des gènes étudiés.

Elle doit aussi se comprendre en fonction de la robustesse des données analysées, de la stratégie d'analyse, des techniques utilisées et du degré de confiance accordé à chaque résultat.

V- BIBLIOGRAPHIE

1. Fitch W.M., Bush R.M., Bender C.A., Cox N.J. ~ Long term trends in the evolution of H(3) HA1 human influenza type A. *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**:7712-7718.
2. Horimoto T., Ito T., Alexander D.J., Kawaoka Y. ~ Cleavability of hemagglutinin from an extremely virulent strain of avian influenza virus containing a unique cleavage site sequence. *Journal of Veterinary Medical Science*, 1995, **57**:927-930.
3. Horimoto T., Rivera E., Pearson J., Senne D., Krauss S., Kawaoka Y., Webster R.G. ~ Origin and molecular changes associated with emergence of a highly pathogenic H5N2 influenza virus in Mexico. *Virology*, 1995, **213**:223-230
4. Subbarao K., Klimov A., Katz J., Regnery H., Lim W., Hall H., Perdue M., Swayne D., Bender C., Huang J., Hemphill M., Rowe T., Shaw M., Xu X., Fukuda K., Cox N. ~ Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness [see comments]. *Science*, 1998, **279**:393-396.
5. Guillot S., Caro V., Cuervo N., Korotkova E., Combiescu M., Persu A., Aubert-Combiescu A., Delpeyroux F., Crainic R. ~ Natural genetic exchanges between vaccine and wild poliovirus strains in humans. *Journal of Virology*, 2000, **74**:8434-8443.
6. Blanchard A., Ferris S., Chamaret S., Guetard D., Montagnier L. ~ Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *Journal of Virology*, 1998, **72**:4537-4540

