

# Diversité génomique virale : évaluation et conséquences sur la virulence des infections respiratoires

Maxime Pichon  
Rodolphe Pelissier  
Martine Valette  
Bruno Lina  
Laurence Josset

Centre hospitalier universitaire de Lyon,  
Laboratoire de virologie, Institut  
des agents infectieux (IAI) de Lyon,  
Centre national de référence des virus  
respiratoires France Sud, Centre  
de biologie et de pathologie Nord,  
Groupement hospitalier Nord  
F-69317, Lyon cedex 04, France  
Université Claude Bernard Lyon 1,  
Faculté de médecine Lyon Est,  
CIRI, ENS, Inserm U1111  
CNRS UMR5308, équipe Virpath,  
F-69372, Lyon cedex 08, France  
<maxime.pichon01@chu-lyon.fr>

**Résumé.** Au cours de l'infection virale, la réplication d'un virus à ARN génère une population diversifiée par accumulation de mutations ponctuelles dans son génome. L'étude de cette diversité génétique virale au moyen des technologies de séquençage à haut débit (séquençage de nouvelle génération [NGS]) a permis une meilleure compréhension de la pathogenèse de l'infection ainsi que des mécanismes d'action des thérapeutiques antivirales. Cette révolution technologique s'est accompagnée d'un nombre impressionnant de nouvelles études centrées sur la diversité génétique d'une souche virale et ses conséquences sur l'histoire naturelle de l'infection. Dans cette revue, nous aborderons les grands principes de l'évaluation de la diversité génétique virale, avant d'en résumer les principales conséquences physiopathologiques des infections à virus respiratoires. À ce jour, bien qu'aucune étude n'établisse clairement son rôle dans la pathogenèse des formes sévères des infections respiratoires, la diversification virale pourrait avoir un double effet. Elle pourrait être un formidable outil de virulence par production d'un avantage évolutif majeur grâce à la diversification des populations virales, c'est-à-dire les quasi-espèces, présentes chez un hôte ou délétère pour le virus en cas de dépassement d'un seuil d'erreur tolérable de variabilité au niveau de leur séquence, aboutissant à terme à son extinction. Il est donc important de l'étudier dans le cadre d'infections à virus respiratoires, comme les virus influenza, les virus respiratoires syncytiaux (VRS) ou les rhinovirus. La compréhension fine de cette propriété permettra d'en envisager de multiples applications cliniques, que ce soit organisationnelles (orientation précoce des patients vers une unité de soins intensifs), thérapeutiques (par le recours à des molécules mutagènes) ou préventives (stabilisation des souches vaccinales pour une plus grande innocuité).

**Mots clés :** diversité virale, virulence, infection respiratoire, sévérité clinique

**Abstract.** During host infection, viral replication generates multiple subpopulations. Studies of viral diversity using high-throughput sequencing technologies provide a better understanding of the therapeutic effects as well as of the viral pathogenesis. This technical evolution led to an impressive number of studies analyzing this viral characteristic. In this review, we will discuss the principles of the evaluation of viral diversity, before summarizing the main physiological consequences for respiratory viruses. To date, although no study clearly established its role in pathogenesis of severe forms, viral diversification can be alternately a formidable virulence advantage or deleterious to the virus, resulting in its extinction (error-threshold). Because of these differences, it is important to study it in the context of respiratory virus infection, such as Influenza, respiratory syncytial virus (RVS) or rhinovirus. The precise understanding of this property allows us to consider multiple clinical applications, *i.e.* therapeutic or preventive.

**Key words:** viral diversity, virulence, respiratory infection, clinical outcome

**Tirés à part :** M. Pichon

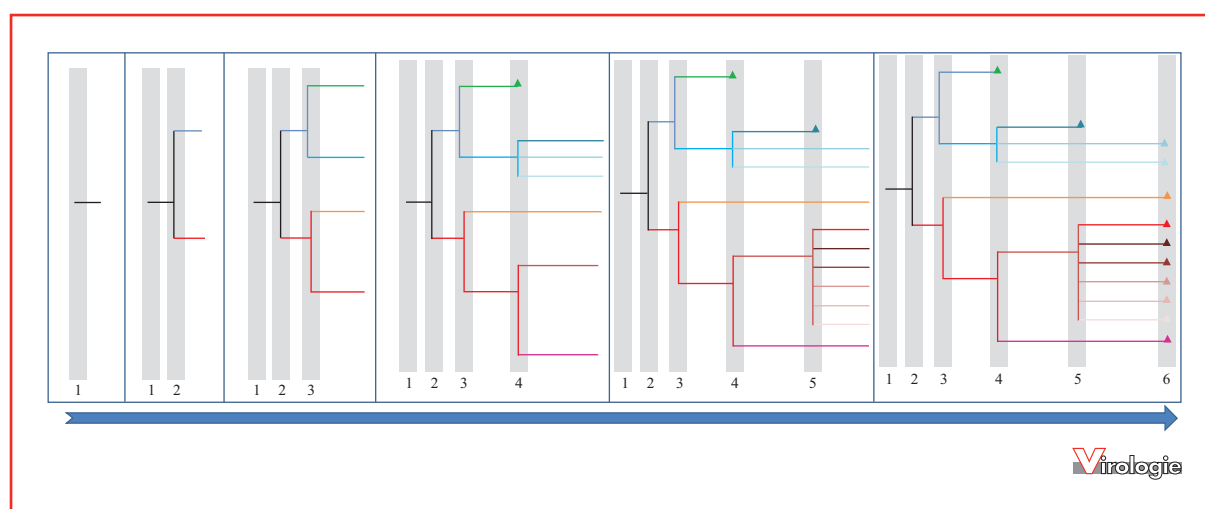
## Introduction

Au cours d'une infection respiratoire impliquant un virus à ARN, la réplication virale génère une population génétiquement diversifiée au sein du tractus respiratoire de l'hôte. Cette diversification est fonction de paramètres liés au virus (nombre de génomes en présence, efficacité et fidélité de la polymérase, modalité de transmission interhôte) mais aussi de l'intégrité de l'hôte infecté (efficacité des barrières physiques ou de la pression de sélection immunitaire).

Chez un hôte, les virus sont présents sous forme de nombreuses populations distinctes, bien que proches, aussi bien en termes de séquences génétiques que de propriétés antigéniques [1-4]. Les variations nucléotidiques s'expliquent par un défaut de fidélité des enzymes répliquatives (polymérases) des virus à ARN ( $10^{-5}$  à  $10^{-6}$  substitutions par nucléotides par cycle réplcatif) comme des virus à ADN ( $10^{-7}$  à  $10^{-8}$  substitutions par nucléotides par cycle réplcatif) [5]. En comparaison, le taux d'erreur d'une polymérase cellulaire est proche de  $10^{-8}$  substitutions par nucléotide par cycle réplcatif, taux limité par le système d'autocorrection de ces enzymes excisant 90 à 99,9 % des anomalies générées par leur activité exonucléase [6]. Le concept de « quasi-espèces », initialement proposé sur des bases purement théoriques par Eigen en 1971, décrit un groupe d'individus différents mais reliés génétiquement au cours de l'apparition de la vie sur terre. Ce concept sous-tend que

les erreurs des polymérases virales, induisant la création de nouveaux mutants isolés ou présents au sein de véritables populations, sont un moteur de l'adaptation des virus à de nouvelles conditions environnementales [7]. Aboutissant parfois à de nombreuses quasi-espèces, ces variants peuvent présenter différents phénotypes pouvant se transmettre d'hôte en hôte au cours de l'étape d'invasion ou être issus de la dissémination de l'infection au sein de différents tissus chez un hôte donné [5, 8-14] (*figure 1*). La variabilité et la plasticité des génomes sont donc des propriétés évolutives inhérentes aux virus (principalement ceux à ARN). La séquence consensus, à distinguer de la séquence majoritaire, retrouvée par les technologies de séquençage classiques ne peut expliquer, à elle seule, toutes les caractéristiques de cette plasticité. En effet, la séquence consensus, reconstruction parfois artificielle, constituée des bases les plus fréquentes à chaque position nucléotidique, ne présume nullement de sa représentativité biologique. Elle est à distinguer de la séquence majoritaire, représentant, comme son nom l'indique, la majorité des séquences retrouvées. Une approche permettant de reconstituer les haplotypes viraux pourrait mettre en évidence les génomes constitués de la même séquence nucléotidique, bien que minoritaires (*figure 2*).

La diversification génomique, ainsi que les quasi-espèces virales qui en résultent, est une notion théorique étudiée depuis près de 40 ans. La démocratisation des technologies



**Figure 1. Représentation de la diversification génomique virale au cours des cycles réplicatifs successifs.** Les erreurs produites à chaque cycle réplcatif (chaque bande grisée représentant un cycle réplcatif) entraînent la divergence des propriétés antigéniques et nucléotidiques des virus (indiquées par des couleurs différentes selon leur divergence), générés et observés après séquençage créant un « spectre de mutants » appelé également « quasi-espèce ». Les triangles représentent la disparition (possible extinction spontanée) du mutant considéré au cours de la réplication virale. Ainsi, l'ensemble de la population virale générée n'est pas forcément observé, en fonction du délai entre la génération du mutant et le séquençage (l'échelle de temps étant représentée par la flèche bleue). Par exemple, un séquençage ayant lieu après le cycle 6 ne verra pas les deux premiers virus (triangles vert et bleu), étant conduits à l'extinction par génération d'une erreur létale.



**Figure 2. Distinction entre quasi-espèce virale, séquence consensus et séquence majoritaire.** Les lignes horizontales ( $n = 10$ ) représentent les génomes des virus présents dans un même échantillon constituant une quasi-espèce. Les ronds, rectangles, triangles ou étoiles représentent des mutations ponctuelles, présentes sur le génome viral du virus séquencé. On constate que la séquence consensus, portant trois mutations (présentes individuellement dans plus de 50 % des virus séquencés) est une reconstruction artificielle n'existant pas parmi la population virale séquencée. La séquence majoritaire correspond à la séquence la plus fréquemment détectée parmi les virus séquencés ( $n = 3$ , < 50 % des virus séquencés). Les séquences consensus, étudiées par les technologies de séquençage classiques, ne permettent donc pas de visualiser l'intégralité des informations portées par les variants minoritaires (bien que leur détection ait été considérablement améliorée depuis l'émergence des NGS comparativement au séquençage de Sanger).

de séquençage à haut débit permet de générer, à moindre coût, des quantités considérables de données portant sur les génomes viraux. Ces données permettent de faire basculer les études théoriques vers de l'analyse expérimentale, ouvrant de vastes possibilités d'investigation aussi bien au niveau de nouvelles thérapeutiques anti-infectieuses que de méthodes de vaccinations innovantes. Se concentrant initialement dans le domaine des virus des hépatites et du VIH, les données produites par des technologies de séquençage du génome complet à haut débit ont permis les études sur la quasi-espèce virale et la diversité virale, tant de virus à ARN qu'à ADN. Le nombre de publications portant sur la diversité virale (mot-clé : « *viral diversity* ») est passé de 321 publications en 2003, avant l'émergence du NGS, à plus de 1000 en 2015 (figure 3). Le nombre de publications portant sur la diversité virale inter- et intrahôte augmente ainsi de façon quasi exponentielle, leur nombre accompagnant le perfectionnement des technologies de séquençage et d'analyse des séquences obtenues.

L'étude de la diversification génomique reste néanmoins un champ d'étude complexe. Les infections respiratoires – virus grippal, virus respiratoire syncytial (VRS) ou rhinovirus (HRV) – étant principalement aiguës, l'importance de la diversification au cours de l'évolution clinique reste imparfaitement comprise. *A contrario* d'une infection chronique, aucun prélèvement n'est généralement réalisé pour contrôler une infection ou suivre la décroissance de celle-ci. Peu d'études ont donc exploré la diversification génomique

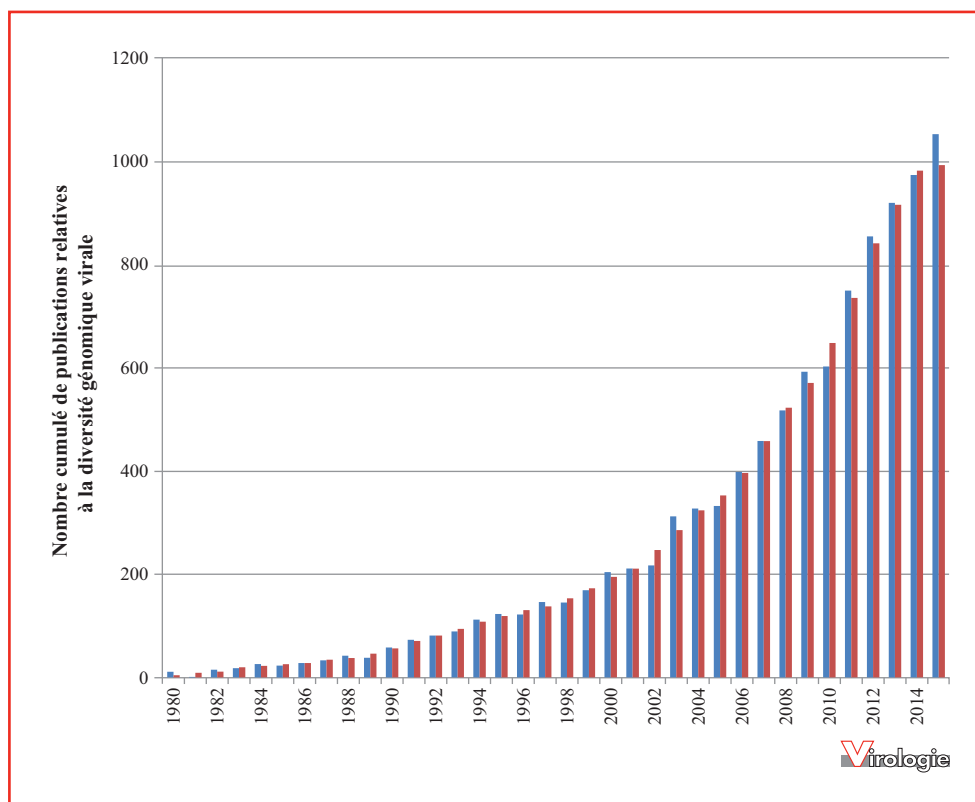
virale au cours de l'infection respiratoire, hormis dans des situations particulières comme chez les patients immunodéprimés à l'origine d'excrétions prolongées [15].

Dans cette revue, nous allons décrire, après en avoir vu les principaux modes d'estimation et méthodes d'analyse informatique, les conséquences cliniques de cette diversification virale dans le cadre de la virulence des infections virales respiratoires chez l'homme.

## Mesure de la diversité génomique

De nombreux paramètres ayant un impact sur la diversité génomique virale (ainsi que sur le *fitness* viral correspondant), il est nécessaire de définir l'unité de mesure permettant de la quantifier (tableau 1).

L'analyse de la diversité génomique virale est fortement impactée par chacune des étapes permettant la génération des données de séquençage (wet-lab), depuis la RT-PCR (taux d'erreur :  $1,8 \times 10^{-4}$ /cycle de PCR), la constitution de la librairie (fragmentation-réparation et indexation) et le séquençage en lui-même ; le taux d'erreur global étant important avant l'analyse, la correction informatique (par élimination des données de qualité insuffisante) le ramène à 0,11 % pour la plateforme HiSeq (Illumina, CA, États-Unis), 0,5 % pour le Ion PGM (LifeTechnologies, CA, États-Unis) ou 15 % pour le PacBio (Pacific Biosciences,



**Figure 3. Données bibliométriques relevant les publications portant sur la diversité virale de 1980 à 2015.** L'échelle verticale représente le total cumulé d'études, répondant à la recherche sur le moteur de recherche PubMed, portant sur la diversité génomique virale et l'échelle horizontale l'année de publication. Les données en bleu représentent le nombre de publications de l'année  $n$  et les données en rouges correspondent à la moyenne du nombre de publications incluant l'année  $n-1$  et  $n+1$ . Le nombre de publications présente une augmentation notable au cours des années 2003 à 2014, correspondant à l'émergence de nouvelles technologies de séquençage.

CA, États-Unis) [16, 17]. Il est donc fondamental de valider le processus analytique par l'utilisation d'un contrôle de qualité adapté à la proportion de diversité artificielle connue. L'utilisation de plasmides, comme ceux utilisés dans les expérimentations de génomique inverse, ou de virus, dits « de plage de lyse », permet de produire des contrôles de qualité satisfaisants [18]. Ceux-ci permettront de valider la fiabilité des étapes de constitution des librairies et des processus analytiques. Les plasmides sont utilisés comme contrôles pour estimer les erreurs de séquençage car ils sont, en principe, strictement identiques avant leur introduction dans le processus analytique. La mise en évidence de variants dans les séquences produites correspondant à ces plasmides permettra donc de mettre en évidence les erreurs de séquençage « artificielles », et ainsi d'estimer la potentielle surestimation de la diversité génomique des virus séquencés au cours de la même expérimentation. Cette manipulation devra néanmoins s'accompagner d'importantes précautions, l'utilisation de plasmides étant parfois source de contamination des zones de laboratoire. Il est important de valider le processus d'interprétation de

données générées de cette manière afin de ne pas introduire de biais de sous- ou de sur-représentation de la diversité génomique [19]. Une surestimation du bruit de fond, à limiter par l'utilisation de standard (échantillon biologique quantifié) ou par l'utilisation de suite logiciel adapté (test des paramètres de filtration au préalable), risque en effet de conduire à une sous-estimation de la diversité des échantillons testés. Réciproquement, l'utilisation de données aux paramètres de qualité non respectés, ainsi que l'utilisation d'un filtre informatique trop peu sévère, peut conduire à une surestimation de la diversité des échantillons testés.

Bien qu'il soit souvent impossible d'identifier toutes les erreurs produites par des étapes précédant le séquençage, que seul un protocole optimisé de réalisation des banques permet de limiter, l'utilisation de logiciels adaptés pour la détection des variants de faible fréquence est nécessaire afin de mesurer finement l'impact des erreurs produites par le séquençage [19]. L'utilisation de ces logiciels permet de s'affranchir des biais *inter-run* plateforme-dépendant en générant des données de qualité satisfaisante même sur les génomes imparfaitement séquencés ou présentant une

**Tableau 1 Principaux estimateurs mathématiques utilisés dans l'étude de la diversité génomique.**

Nom	Abr.	Formule	Groupe	Avantages	Inconvénients
Fréquences minimales maximales	$n$	$n = \sum_{k=1}^{S_{obs}} fk$	Abondance fonctionnelle	Simplicité d'utilisation	Population de taille finie
Diversité nucléotidique	$\pi$	$\pi = Nu/L$	Abondance fonctionnelle	Simplicité d'utilisation	Population de taille finie
Indice de Simpson (1949)	$\lambda$	$\lambda = \sum_{s=1}^S p^2$	Abondance relative (indice $\alpha$ )	Donne des valeurs plus intuitives	Donne plus de poids aux espèces abondantes
Indice de Shannon (1949)	$H$	$H(p) = \sum_{i=1}^K pi \log(pi)$	Abondance relative (indice $\alpha$ )	Insensible à la richesse spécifique	Saturabilité
Indice de Hill (1973)	$N_h$	$N_h(\alpha) = \left(\sum_{i=1}^S pi\right)^{1/(1-\alpha)}$	Abondance relative (indice $\alpha$ )	Synthèse des indices de Shannon et de Simpson	Interprétation difficile en l'absence de l'évaluation de l'indice de Shannon ou de Simpson
Indice de Chao 2 (1987)	$S$	$S = D + \frac{(N-1)\binom{f^2_1}{2}}{2N(f_2)}$	Abondance relative (indice $\alpha$ )	Calcule le nombre d'espèces non observées à partir de celles observées	Suppose que $N$ est grand et probabilité d'observation de chaque espèce est petite. Ne donne qu'un nombre minimum d'espèces

Abr : abréviation ;  $p$  : la fréquence relative des espèces ;  $N_u$  : nombre moyen de différences nucléotidiques entre paires de séquences ;  $L$  : nombre total de sites ;  $N$  : nombre d'individus ;  $D$  : espèces différentes parmi les  $S$  existantes. L'année de publication de l'indice est indiquée entre parenthèses.

importante variabilité comme ceux des virus influenza. La quantification des variants de faible fréquence, constituant la majorité de la diversité génomique virale, est donc une variable pouvant engendrer d'importants biais d'analyse, *via* une élévation artificielle de la diversité virale [20]. L'analyse des variants de faible fréquence, reflet de la diversité génétique virale, doit donc prendre en compte les biais induits par les analyses informatiques afin de conclure de manière prudente et raisonnée sur leur impact biologique. Une étude récente de Debbink *et al.*, parue en 2017, insiste, par ailleurs, sur l'intérêt d'une standardisation de la charge virale préalablement à la constitution de la librairie. D'importantes variations de charge virale dans l'échantillon influencent la spécificité et la sensibilité de la détection des polymorphismes génomiques et, donc, la quantité de diversité génomique mise en évidence [21]. Enfin, il est important d'insister sur la nécessité d'analyser, autant que possible, les haplotypes viraux, à savoir le

génom viral dans son intégralité. La reconstitution d'un génome de plusieurs kilobases, morcelé en inserts de tailles variables (en fonction de la technologie employée), est nécessaire dans le cadre de séquençages réalisés avec les technologies en émulsion (LifeTechnologies, CA, États-Unis) ou sur phase solide (Illumina, San Diego, États-Unis) [22]. Ces technologies, dites de deuxième génération, permettent le séquençage en grande quantité de courtes séquences, de 75 à 300 bases obtenues après fragmentation (enzymatique ou mécanique). Un assemblage *in silico* des génomes présents dans l'échantillon séquençé est donc nécessaire pour permettre de relier entre elles les mutations observées et d'en connaître précisément les proportions (figure 2). Des solutions informatiques peuvent être appliquées pour les reconstituer. Les logiciels utilisés à cette fin restent néanmoins imparfaits, basés sur des reconstructions probabilistes. Ces algorithmes de travail sont donc parfois à l'origine de diversification artificiellement augmentée ou

diminuée, permettant d'estimer (sans les définir précisément) le nombre d'haplotypes. Pour limiter ces problèmes, l'utilisation d'une technologie de troisième génération comme le PacBio (Pacific Biosciences, CA, États-Unis) ou le MinIon (Nanopore Technologies, Oxford, Royaume-Uni) pourrait être envisagée. Elle permet de séquencer les acides nucléiques en un seul « *read* », sans avoir besoin préalablement de fragmenter les génomes viraux. Il serait donc possible de connaître la constitution complète des génomes analysés afin de déterminer la proportion réelle de chaque séquence (figure 2). La diversité réelle en serait donc plus finement quantifiée. La limite importante induite par ces technologies, les rendant inutilisables en pratique pour la détermination des quasi-espèces au niveau haplotypique, reste leur taux considérable d'erreur (15 à 20 % d'erreur pour le PacBio avant traitement informatique). Cela doit donc être encore amélioré pour en permettre l'utilisation dans cette problématique [23].

## Conséquences phénotypiques de la diversification génomique virale sur la virulence des infections virales

### *La diversification génomique virale comme facteur de virulence*

Deux études reliant diversification génomique virale et virulence ont été réalisées sur les poliovirus [24, 25].

Après avoir mis en évidence une mutation (G64S) conférant une haute fidélité à la polymérase virale sans en altérer l'efficacité répliquative, Vignuzzi *et al.* ont constaté que la répllication virale des virus mutés générait une population à faible diversité génétique et virulence. *In vivo*, une telle population clonale portant cette mutation G64S ne pouvait être létale qu'administrée à très haute concentration et directement dans le système nerveux central (SNC). En effet, les virus mutés G64S ne pouvaient pas passer la barrière hématoencéphalique, le SNC étant le site majeur de répllication de ce virus. Afin de s'assurer que l'absence de virulence constatée n'était pas due à la seule présence de cette mutation, le virus G64S a été artificiellement diversifié, par adjonction de ribavirine et de 5-fluoro-uracile au milieu de culture diminuant ainsi temporairement la fidélité de la polymérase virale [26]. La nouvelle population virale générée présentait une diversité génomique importante et la même dose létale pour la souris que les souches virales sauvages (1,6 millions de plaque forming-unit (PFU) (unité formant plaque). Enfin, les auteurs ont montré que cette différence de létalité n'était pas due à l'émergence d'un clone hautement létal. Moins diversifiés (deux à trois fois moins de mutations), les virions retrouvés par prélèvement au

niveau du SNC ne présentaient pas la virulence de la population virale dans son ensemble. La réinjection à distance du SNC de ces seuls virions à une souris saine ne conduisait pas à une dissémination neuroméningée de ces virus, justifiant leur faible létalité. Cette preuve expérimentale est schématisée par la figure 4.

Ces résultats ont conforté l'idée que les quasi-espèces générées par la diversification génomique étaient nécessaires à la dissémination méningée de ce virus. Les auteurs ont avancé l'hypothèse d'une coopération entre variants viraux au cours du franchissement des multiples barrières rencontrées au cours de l'infection ; certains variants facilitant la colonisation et le franchissement de l'épithélium intestinal, d'autres servant de leurre pour tromper le système immunitaire alors que les derniers, altérant la barrière hématoencéphalique, permettent son franchissement.

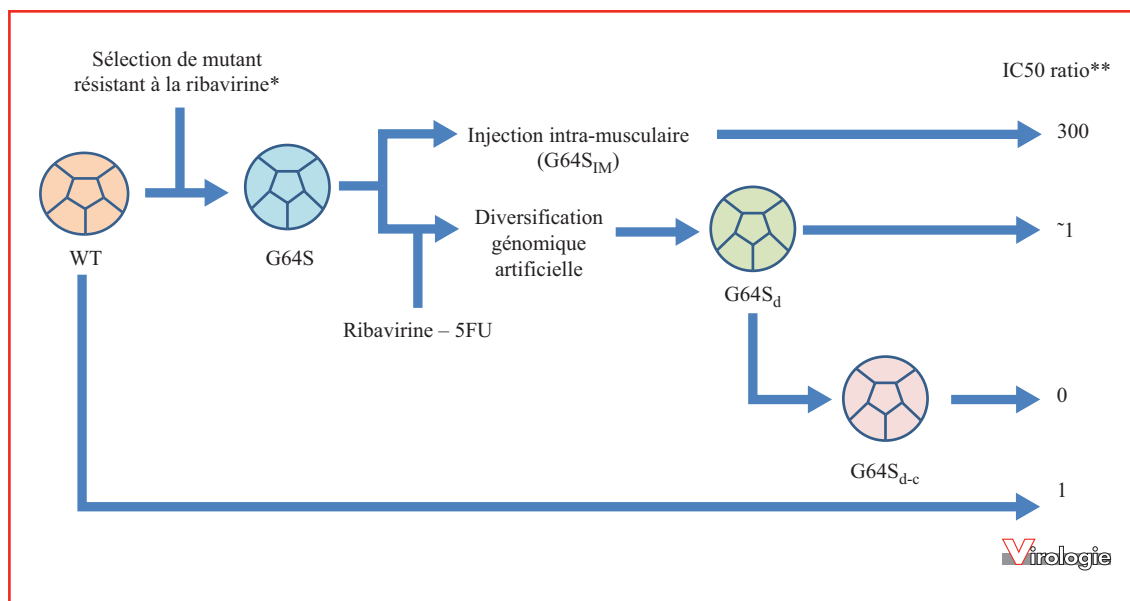
### *La diversification génomique virale limitant la virulence*

Comme cela fut prouvé dans le modèle poliovirus, la ribavirine, molécule antivirale efficace contre de nombreuses cibles virales (*Herpesviridae*, Orthomyxovirus, Paramyxovirus, Filovirus. . .), peut aussi augmenter la fréquence de mutation au-delà d'un seuil tolérable d'erreur. Le franchissement de ce seuil amènerait le virus à une extinction virale [26, 27]. Contrairement aux modèles d'entérovirus et de poliovirus pour lesquels la diversification augmenterait la virulence jusqu'à un certain seuil de tolérance, la ribavirine pourrait donc provoquer une véritable extinction virale du fait d'une importante mutagenèse pour certaines espèces virales telles que le virus de la stomatite vésiculeuse [VSV] et les hantavirus, par exemple [28-32]. Cette notion, symbolisant le dépassement du seuil d'erreur toléré, est appelée *error catastrophe* [33].

La première démonstration de l'aspect délétère de l'accumulation de mutations dans un génome viral a été réalisée par Holland *et al.* dans un modèle animal d'infection à VSV [34]. Ces auteurs ont pu démontrer qu'après exposition à divers mutagènes viraux (5-fluoro-uracile, 5-azacytidine), les virus retrouvés infectieux présentaient un taux de mutation identique à celui des virus sauvages. Ces résultats confortent l'idée d'un seuil de mutations au-delà duquel la répllication virale chute drastiquement, expliquant la disparition des virus comportant un nombre de mutations supérieur à ce seuil.

La présence d'une sous-population virale artificiellement diversifiée par l'action de molécules mutagènes pourrait aussi entraver le pouvoir infectieux de la sous-population virale « naturelle ». Ce phénomène, parfaitement étudié dans le cas du syndrome pied-main-bouche, mais imparfaitement compris dans d'autres infections virales, est dénommé « défection létale » [35, 36]. La principale





**Figure 4. Représentation de l'impact clinique de la diversification génomique virale dans un modèle murin d'infection à Poliovirus (d'après [24]).** IC50 : quantité suffisante pour tuer 50 % de la population murine expérimentale ; WT : poliovirus sauvage ; G64S : poliovirus mutant portant la mutation G64S augmentant la fidélité de la polymérase virale ; G64S<sub>d</sub> : population de poliovirus mutant G64S diversifiée artificiellement ; G64S<sub>d-c</sub> : population de poliovirus mutant G64S diversifié artificiellement injectée puis récupérée dans le liquide cébrospinal murin ; \* : sélection de mutant par culture ; \*\* : ratio déterminé par normalisation de l'IC50 à l'IC50 du virus sauvage.

hypothèse mécanistique se baserait sur la formation de complexes protéiques inefficaces codés par les virus porteurs d'un nombre anormalement élevé de mutations. Ces complexes limiteraient la réplication de la sous-population sauvage entraînant ainsi une diminution de la charge virale globale. Une étude de 2011 a montré, dans un essai clinique portant sur le VIH-1, l'applicabilité clinique des mutagenèses virales létales [37]. Le taux de mutation du VIH-1 étant proche du seuil d'erreur toléré, l'augmentation, même minime, de ce taux de mutation conduit à l'extinction rapide de la population virale. Il reste néanmoins impossible d'extrapoler les résultats de test *in vitro* afin de prédire le délai et l'intensité de l'effet de telles thérapeutiques appliquées aux infections humaines, du fait des réservoirs en présence, des variations dans le *fitness* viral et de l'intensité de la réponse immunitaire de l'hôte infecté.

De même, les molécules habituellement utilisées dans le traitement de certaines néoplasies, comme le 5-fluorouracile, ont prouvé leur efficacité dans les modèles murins d'infection par les arénavirus [38]. Néanmoins, aucune application médicale n'a été validée à ce jour, les niveaux de mutagenèse nécessaires pour une application d'intérêt médical variant énormément entre les virus et les situations cliniques. De plus, ces seuils varient en fonction de la région nucléotidique porteuse de diversité, de la thérapeutique considérée ainsi que de la génétique du patient. La variabilité des cas à étudier est telle que l'utilisation de

ces notions en pratique clinique courante en est impossible pour le moment.

En conclusion, il apparaît que le taux de mutation viral est finement déterminé durant l'évolution d'un virus. Alors qu'une accumulation de mutations peut mener à l'extinction d'une population virale, leur absence ou leur nombre insuffisant limite la « survie » en environnement changeant, en cas de conditions hostiles comme la réponse immunitaire de l'hôte. La « survie » d'un virus donné est donc la résultante d'un équilibre entre la plasticité et la stabilité génomique, facilitant l'adaptation à de nouvelles niches écologiques mais conservant intact le pouvoir réplcatif du virus.

## La diversification génomique des virus respiratoires

Contrairement aux modèles présentés ci-dessus, aucune étude à ce jour n'a pu mettre en évidence de lien entre la diversité génomique des virus respiratoires et la sévérité clinique de la maladie dont ils sont responsables. Les virus respiratoires, principalement à ARN (influenza, VRS et HRV), ont néanmoins fait l'objet de nombreuses publications récentes.

Avec un taux de mutation estimé entre  $7,1 \times 10^{-6}$  à  $4,5 \times 10^{-5}$  substitutions par nucléotide par cycle d'infection

cellulaire, la faible fidélité de la polymérase du virus influenza est le moteur de son évolution [5]. De plus, une étude récente a montré que le taux de mutations du virus influenza, évalué par tests de fluctuation, était deux fois supérieur à celui constaté par NGS [39, 40]. Dans cette méthode, un nombre important de cultures en parallèle étaient infectées avec un inoculum faible, permettant l'émergence de mutants résistants après une période de croissance exponentielle. Les mutations étant des événements rares et aléatoires, elles suivent un modèle mathématique dit de « Poisson ». Utilisant la résistance aux anticorps monoclonaux comme mode de sélection, cette méthode permet une estimation fiable du taux de mutation par cycle réplcatif. Ces mutations confirmant la diversité génétique importante des souches de grippe, concernent principalement les transversions de A vers G et de U vers C. Chaque cycle réplcatif induisant deux à trois mutations et 28 à 31 % des mutations ayant une conséquence « létale » sur le virion, la probabilité qu'un génome viral muté se réplique efficacement n'est que de 34 à 49 % [41]. Après génération de 28 à 31 % de mutations directement non viables pour les virions produits, les auteurs estiment que près de 99 % des virions produits sont non infectieux, indépendamment du modèle cellulaire infecté, plaçant les virus influenza proches de la limite du taux d'erreur tolérable par une souche virale pour rester réplcative [39]. La génération de souches non infectieuses peut présenter un impact important car il a été montré que les particules virales défectives pourraient avoir un impact sur la virulence de l'infection. La production de ces particules au cours de la réplcation virale est ainsi un facteur de limitation potentiel de la virulence de cette infection [42].

L'étude de la diversification intra- et inter-hôte des génomes viraux est donc un élément central de la compréhension de la pathogenèse virale, aussi bien dans sa genèse que dans ses interactions avec le système immunitaire de l'hôte. Cette compréhension est fondamentale pour construire des modèles évolutifs de la grippe [10, 43]. La diversité génomique du virus de la grippe chez l'homme a été surtout étudiée chez les patients immunodéprimés et dans le cadre des résistances aux antiviraux [44]. Il est intéressant de noter que les antiviraux mutagènes ne génèrent pas de résistance, même après de nombreux passages en culture cellulaire [42]. Néanmoins, en l'absence de données cliniques fiables associées, les caractéristiques de la diversité génomique grippale, au temps précoce de la maladie, restent encore incompréhensibles chez les patients au système immunitaire efficace, même si les données disponibles aujourd'hui suggèrent une diversité globalement faible lors d'une infection aiguë.

Pour le VRS, une étude a montré la présence de régions du génome viral hautement (points chauds de mutation) ou faiblement (points froids de mutation) diversifiées au

sein d'une même souche [45]. La région codant la protéine G (protéine de surface fixant les récepteurs cellulaires) étant décrite comme la plus variable du génome des VRS, la diversité virale peut aussi être considérée comme une méthode d'amélioration du *fitness* viral ou de tentative d'échappement au système immunitaire de l'hôte infecté. Par opposition, le gène de la protéine M, protéine structurale au carrefour de nombreuses fonctions virales, est le gène le moins diversifié de l'ensemble du génome du VRS. De façon globale, le nombre important de points froids de mutations conforte l'idée selon laquelle le VRS est un virus restreint dans sa diversification. Cette constatation est cohérente avec la faible diversification épidémiologique décrite au cours des 50 dernières années [46]. Le taux de mutation du VRS reste néanmoins à étayer. Estimé uniquement dans certaines situations épidémiques, ce taux est de  $4,95 \times 10^{-3}$  substitutions par nucléotide par cycle d'infection cellulaire, soit cinq fois plus que ce qui est généralement rapporté pour ce virus. Ce taux est surtout 100 à 1000 fois plus important que celui relevé pour le virus influenza, dont l'évolution phylogénique est pourtant plus rapide [47]. Ces données sont néanmoins à mettre en balance avec l'absence de corrélation entre ces variations génétiques et la sévérité clinique de l'infection.

La transmission inter-hôte peut être aussi un facteur impactant la diversité génomique virale, en amont de toute diversification. Dans le cadre d'une infection expérimentale à HRV, il a été montré que la diversité génomique virale initialement retrouvée chez un hôte nouvellement infecté n'est pas limitée par la transmission préférentielle d'une faible proportion des variants minoritaires. La quasi-totalité des variants est en effet secondairement retrouvée chez les hôtes après infection [48]. Il est par ailleurs intéressant de noter que les auteurs ont pu estimer, pour ce virus, un taux de mutations intermédiaire, entre  $10^{-4}$  et  $10^{-3}$  substitutions par cycle d'infection cellulaire. Au contraire, dans une étude analysant la transmission de la diversité génétique des virus influenza, il a pu être mis en évidence l'association entre la sévérité de la maladie grippale (estimé par la température de l'hôte infecté) et la limitation de la diversité transmise d'un hôte à l'autre. Ainsi, une restriction de la diversité génétique de la souche transmise était associée à une plus grande sévérité chez l'hôte nouvellement infecté [49]. Cette information est cohérente avec l'hypothèse de nécessité d'acquisition d'un large panel de mutations et/ou de réassortiments pour permettre l'obtention d'un *fitness* optimisé chez un nouvel hôte [13].

La diversification doit donc se voir comme la résultante de la proportion (et du taux respectif de mutations) de ces régions génomiques chaudes et froides. Cette diversification théorique reste néanmoins à corrélérer à la possibilité de transmission de ces variants minoritaires [13]. Il faut donc distinguer diversification à court terme et diversification à



long terme. Cette diversité à court terme est toujours supérieure à la diversité à long terme, suggérant la sélection d'une fraction des variants générés [47].

## Conclusion

La prodigieuse évolution des techniques de biologie moléculaire, avec, au premier plan, les techniques de séquençage à haut débit dites de « nouvelles générations », ouvre de nouveaux champs d'étude en microbiologie. L'étude du génome viral, initié par le séquençage selon la méthode de Sanger, est ainsi grandement améliorée. De nombreux travaux suggèrent qu'il existe un lien fort entre la diversification génomique du pathogène et la sévérité clinique de l'infection, au stade initial ou au cours de la maladie. Il reste néanmoins difficile d'obtenir des données cliniques fiables dans le cadre des infections respiratoires. En effet, les formes bénignes de ces maladies ne sont pas systématiquement relevées par les études épidémiologiques au contraire des formes sévères qui sont celles principalement étudiées. Il sera donc nécessaire de mettre en place des études prospectives, aux prélèvements itératifs systématiques, pour obtenir des données extrapolables. La compréhension fine du lien entre la diversité génomique virale et l'histoire naturelle de la maladie permettrait d'améliorer la compréhension de la pathophysiologie des maladies infectieuses et d'identifier des biomarqueurs utilisables en routine. Cette amélioration permettrait ainsi d'optimiser la prise en charge en service spécialisé des patients présentant des infections respiratoires aux complications prévisibles. La détermination des espèces virales sensibles au seuil d'*error catastrophe* permettra d'envisager le développement de thérapies innovantes, déstabilisant la polymérase virale, augmentant ainsi son taux d'erreur et conduisant à l'extinction de la souche. Enfin, l'identification des espèces virales aux pathogénicités positivement corrélées au taux de mutation pourrait permettre d'envisager la production de souches vaccinales atténuées présentant une sécurité d'utilisation augmentée car disposant d'une polymérase de fidélité améliorée. L'étude de la diversité a donc des applications cliniques importantes. Néanmoins, les limites de ces études restent principalement organisationnelles et financières. À terme, elles permettront néanmoins de proposer des modifications majeures pour la prise en charge des maladies infectieuses virales.

**Liens d'intérêts :** les auteurs déclarent ne pas avoir de lien d'intérêt en rapport avec cet article.

## Références

- Ghedini E, Fitch A, Boyne A, *et al.* Mixed infection and the genesis of influenza virus diversity. *J Virol* 2009 ; 83 : 8832-41.
- Ghedini E, Laplante J, DePasse J, *et al.* Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J Infect Dis* 2011 ; 203 : 168-74.
- Ghedini E, Holmes EC, DePasse JV, *et al.* Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. *J Infect Dis* 2012 ; 206 : 1504-11.
- McCrone JT, Luring AS. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. *J Virol* 2016 ; 90 : 6884-95.
- Sanjuan R, Nebot MR, Chirico N, *et al.* Viral Mutation Rates. *J Virol* 2010 ; 84 : 9733-48.
- Kunkel TA. DNA Replication Fidelity. *J Biol Chem* 2004 ; 279 : 16895-8.
- Eigen M. Molecular self-organization and the early stages of evolution. *Experientia* 1971 ; 27.
- Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev* 2012 ; 76 : 159-216.
- Poon LLM, Song T, Rosenfeld R, *et al.* Quantifying influenza virus diversity and transmission in humans. *Nat Genet* 2016 ; 48 : 195-200.
- Murcia PR, Hughes J, Battista P, *et al.* Evolution of an Eurasian avian-like influenza virus in naïve and vaccinated pigs. *PLoS Pathog* 2012 ; 8 : 1002730.
- Hughes J, Allen RC, Baguelin M, *et al.* Transmission of equine influenza virus during an outbreak is characterized by frequent mixed infections and loose transmission bottlenecks. *PLoS Pathog* 2012 ; 8 : e1003081.
- Wilker PR, Dinis JM, Starrett G, *et al.* Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nat Commun* 2013 ; 4.
- Varble A, Albrecht RA, Backes S, *et al.* Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe* 2014 ; 16 : 691-700.
- Parvin JD, Moscona A, Pan WT, Leider JM, Palese P. Measurement of the mutation rates of animal viruses : influenza A virus and poliovirus type 1. *J Virol* 1986 ; 59.
- Grad YH, Newman R, Zody M, *et al.* Within-host whole-genome deep sequencing and diversity analysis of human respiratory syncytial virus infection reveals dynamics of genomic diversity in the absence and presence of immune pressure. *J Virol* 2014 ; 88 : 7286-93.
- Zhao C, Liu F, Pyle AM. An ultra-processive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* 2018 ; 24 : 183-95.
- Song L, Huang W, Kang J, *et al.* Comparison of error correction algorithms for Ion Torrent PGM data : application to hepatitis B virus. *Sci Rep* 2017 ; 7 : 8106.
- Hoffmann E, Krauss S, Perez D, Webby R, Webster RG. Eight-plasmid system for rapid generation of influenza virus vaccines. *Vaccine* 2002 ; 20 : 3165-70.
- Wilm A, Aw PPK, Bertrand D, *et al.* LoFreq : a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res* 2012 ; 40 : 11189-201.
- Acevedo A, Andino R. Library preparation for highly accurate population sequencing of RNA viruses. *Nat Protoc* 2014 ; 9 : 1760-9.
- Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, *et al.* Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathog* 2017 ; 13 : e1006194.
- Malhotra R, Prabhakara S, Poss M, *et al.* *Estimating Viral Haplotypes in a Population Using k-mer Counting*. Berlin : Heidelberg, 2013.
- Thompson JF, Milos PM. The properties and applications of single-molecule DNA sequencing. *Genome Biol* 2011 ; 12 : 217.
- Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 2006 ; 439.

25. Pfeiffer JK, Kirkegaard K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* 2005 ; 1 : e11.
26. Crotty S, Maag D, Arnold JJ, *et al.* The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nat Med* 2000 ; 6 : 1375-9.
27. Crotty S, Cameron CE, Andino R. RNA virus error catastrophe : direct molecular test by using ribavirin. *PNAS USA* 2001 ; 98 : 6895-900.
28. Airaksinen A, Pariente N, Menéndez-Arias L, Domingo E. Curing of foot-and-mouth disease virus from persistently infected cells by ribavirin involves enhanced mutagenesis. *Virology* 2003 ; 311 : 339-49.
29. Day CW, Smee DF, Julander JG, *et al.* Error-prone replication of West Nile virus caused by ribavirin. *Antiviral Res* 2005 ; 67 : 38-45.
30. Maag D, Castro C, Hong Z, Cameron CE. Hepatitis C virus RNA-dependent RNA polymerase (NS5B) as a mediator of the antiviral activity of ribavirin. *J Biol Chem* 2001 ; 276 : 46094-8.
31. Lanford RE, Chavez D, Guerra B, *et al.* Ribavirin induces error-prone replication of GB virus B in primary tamarin hepatocytes. *J Virol* 2001 ; 75 : 8074-81.
32. Severson WE, Schmaljohn CS, Javadian A, Jonsson CB. Ribavirin causes error catastrophe during Hantaan virus replication. *J Virol* 2003 ; 77 : 481-8.
33. Anderson JP, Daifuku R, Loeb LA. Viral error catastrophe by mutagenic nucleosides. *Annu Rev Microbiol* 2004 ; 58 : 183-205.
34. Holland JJ, Domingo E, de la Torre JC, Steinhauer DA. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J Virol* 1990 ; 64 : 3960-2.
35. Grande-Pérez A, Lázaro E, Lowenstein P, Domingo E, Manrubia SC. Suppression of viral infectivity through lethal defection. *PNAS USA* 2005 ; 102 : 4448-52.
36. González-López C, Arias A, Pariente N, Gómez-Mariano G, Domingo E. Preextinction viral RNA can interfere with infectivity. *J Virol* 2004 ; 78 : 3319-24.
37. Mullins JI, Heath L, Hughes JP, *et al.* Mutation of HIV-1 genomes in a clinical population treated with the mutagenic nucleoside KP1461. *PloS One* 2011 ; 6 : e15135.
38. Ruiz-Jarabo CM, Ly C, Domingo E, de la Torre JC. Lethal mutagenesis of the prototypic arenavirus lymphocytic choriomeningitis virus (LCMV). *Virology* 2003 ; 308 : 37-47.
39. Pauly MD, Procario MC, Lauring AS. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *eLife* 2017 ; 6.
40. Luria SE, Delbrück M. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* 1943 ; 28 : 491-511.
41. Visher E, Whitefield SE, McCrone JT, Fitzsimmons W, Lauring AS. The Mutational Robustness of Influenza A Virus. *PLoS Pathog* 2016 ; 12 : e1005856.
42. Pauly MD, Lauring AS. Effective lethal mutagenesis of influenza virus by three nucleoside analogs. *J Virol* 2015 ; 89 : 3584-97.
43. Dinis J, Florek NW, Fatola OO, *et al.* Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *J Virol* 2016 ; 90 : 3355-65.
44. Rogers MB, Song T, Sebra R, *et al.* Intrahost dynamics of antiviral resistance in influenza a virus reflect complex patterns of segment linkage, reassortment, and natural selection. *mBio* 2015 ; 6.
45. Do LA, Wilm A, Van Doorn HR, *et al.* Direct whole-genome deep-sequencing of human respiratory syncytial virus A and B from Vietnamese children identifies distinct patterns of inter- and intra-host evolution. *J Gen Virol* 2015 ; 96 : 3470-83.
46. Tan L, Lemey P, Houspie L, *et al.* Genetic variability among complete human respiratory syncytial virus subgroup A genomes : bridging molecular evolutionary dynamics and epidemiology. *PloS One* 2012 ; 7 : e51439.
47. Agoti CN, Munywoki PK, Phan MVT, *et al.* Transmission patterns and evolution of respiratory syncytial virus in a community outbreak identified by genomic analysis. *Virus Evol* 2017 ; 3.
48. Cordey S, Junier T, Gerlach D, *et al.* Rhinovirus Genome Evolution during Experimental Human Infection. Martin DP, éditeur. *PLoS ONE* 2010 ; 5 : e10588.
49. Sobel Leonard A, Weissman D, Greenbaum B, Ghedin E, Koelle K. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *J Virol* 2017 : JVI.00171-VI.217.