► L'étude des maladies virales humaines, animales et végétales a récemment bénéficié du développement de la métagénomique en écologie. Cette approche, dite sans *a priori*, consiste à décrire le génome viral total, ou virome, d'une fraction d'un écosystème sans tenir compte de cibles moléculaires déjà connues. Les premiers résultats de la métagénomique en pathologie virale ont été obtenus très localement au niveau d'un tissu ou d'un organe. Ils ont débouché sur la découverte de nouveaux virus et ont permis des avancées sur le plan du diagnostic, de l'épidémiologie moléculaire et de l'évolution virale. Ces travaux sont d'une grande pertinence pour réévaluer des concepts en pathologie et, notamment, le rôle biologique des virus dans un organisme. Ils révèlent la nécessité d'actualiser les méthodes de diagnostic. ◀

La métagénomique microbienne et virale : une nouvelle approche en écologie microbienne

La métagénomique (voir *Glossaire*) est un terme qui est apparu en 1998 dans un article écrit par Jo Handelsman *et al.* [1]. Il faut toutefois remonter au début des années 1990 pour trouver trace des premiers travaux de métagénomique [2], et même aux années 1980 pour deviner l'apparition de ce nouveau concept grâce aux travaux de N.R. Pace *et al.* [3] qui proposent pour la première fois de cloner directement les gènes 5S et 16S des ARN ribosomiaux (ARNr) à partir d'échantillons prélevés dans l'environnement. Le terme métagénomique est ensuite peu à peu devenu le terme consacré pour désigner les travaux de génomique environnementale, génomique des communautés, écogénomique, génomique des populations microbiennes, etc. Les études engagées à la fin du xx^e siècle avaient toutes pour objectif de dépasser le cadre des collections de souches de microorganismes des laboratoires en séquençant directement les acides nucléiques collectés *in situ* dans différents écosystèmes

Métagénomique virale et pathologie

Une histoire récente

Pauline Bernardo¹, Emmanuel Albina^{2,3}, Marc Eloit⁴, Philippe Roumagnac¹

¹ Cirad (Centre de coopération internationale en recherche agronomique pour le développement), UMR BGPI (biologie et génétique des interactions plante-parasite), 34398 Montpellier Cedex 5, France ;

² Cirad, UMR CMAEE (contrôle des maladies animales exotiques et émergentes), 97170 Petit-Bourg, Guadeloupe, France ;

³ Inra, UMR1309 CMAEE, 34398 Montpellier, France ;

⁴ Institut Pasteur, laboratoire de découverte de pathogènes, département de virologie, 28, rue du Docteur Roux, 75015 Paris, France. philippe.roumagnac@cirad.fr

[4]. Ce nouvel angle de recherche permettait de s'affranchir de l'isolement et de la culture des souches bactériennes et, *de facto*, d'avoir accès aux acides nucléiques des souches non cultivables, souches qui représentent plus de 99 % de la biodiversité des micro-organismes [5, 40]. Cette nouvelle approche en écologie microbienne visait ainsi à mieux décrypter la diversité des microorganismes présents au sein d'une niche écologique, à mieux comprendre les relations évolutives des différents taxons identifiés et à caractériser de nouveaux gènes et de nouvelles fonctions [4]. Craig Venter et son équipe [6] ont démontré de façon spectaculaire la puissance de cette nouvelle approche en identifiant 148 phylotypes et 1,2 millions de nouveaux gènes par la seule analyse de quelques échantillons prélevés dans la mer des Sargasses. À la suite des travaux pionniers de T. Allander *et al.* [7], la métagénomique virale émerge en 2002 avec un papier fondateur de M. Breitbart, *et al.* [8]. L'approche utilisée, dite sans *a priori*, consiste à purifier l'ensemble des particules virales d'un échantillon d'un écosystème ciblé puis, indépendamment de cibles moléculaires déjà connues (approche indépendante de séquence), à amplifier le génome viral total (virome) en utilisant des amorces aléatoires. L'ensemble de ces études, décrites dans environ une trentaine d'articles publiés fin 2012, met en évidence un nombre considérable de séquences d'ADN inconnues (> 50 %) ne correspondant à aucune donnée actuellement archivée dans les bases de données internationales [9]. Cette « matière noire » représente le plus important réservoir d'information génétique encore non décodée au sein de la biosphère. De plus, quand les séquences en acides aminés peuvent être décodées, elles présentent le plus souvent des similitudes

Vignette (Photo © Jean Kanitakis).

Figure 1. Nombre de publications en métagénomique enregistrées annuellement par le NCBI (national center for biotechnology) depuis 2001. Les recherches ont été faites avec les mots-clés suivants : virus AND metagenomics OR next generation sequencing OR high throughput sequencing.

très faibles (< 50 %) avec des séquences déjà connues et caractérisées [9]. Les virus environnementaux sont plus ou moins proches des virus actuellement connus. Ils sont généralement rattachés à de nouvelles espèces virales qui, à leur tour, permettent de densifier les arbres phylogénétiques des familles ou des genres viraux [10].

L'ensemble de ces premiers résultats de métagénomique virale et microbienne a contribué à remettre en cause plusieurs concepts communément admis en microbiologie.

Le premier concept est d'ordre ontologique. Ces études révèlent la complexité de la diversité globale des génomes présents au sein des écosystèmes naturels, avec nombre de génomes chimériques et des populations à forte hétérogénéité génétique. Les concepts de génomique horizontale [11] et de métaorganismes [12] sont en train d'émerger ; ils se fondent sur l'idée suivante : un métagénome peut être considéré comme une ressource génomique commune, partagée par un ensemble d'entités aux contours plastiques (le métaorganisme) [12]. Le microbiome humain est sûrement le métaorganisme le plus connu : les cellules bactériennes qui le composent sont dix fois plus nombreuses que les cellules du corps humain les hébergeant [13]. Il a été récemment montré que le microbiome du tube digestif humain est le siège d'échanges génétiques, ainsi que de phénomènes de régulations, de mutualisme et de compétition [14]. La question de la définition d'un être pluricellulaire, l'hôte *stricto sensu* ou bien l'hôte et son cortège de microorganismes, a donc été posée pour essayer de mieux comprendre l'adaptation de l'hôte à son environnement [12].

Le deuxième concept que l'on peut interroger est d'ordre à la fois écologique et pathologique. Les microbes au sens large, et les virus en particulier, sont-ils nécessairement dangereux, en particulier pour la santé de leurs hôtes eucaryotes [15] ? Plusieurs études récentes ont montré que des micro-organismes peuvent être, soit parasitaires, soit mutualistes, en fonction des paramètres environnementaux. À titre d'exemple, il a été publié que des staphylocoques de la flore com-

mensale de l'épithélium cutanée jouent un rôle clé dans la modulation de l'inflammation déclenchée par une blessure cutanée [16].

Les interrogations relatives à ces deux concepts, et surtout le nombre croissant de nouveaux microorganismes découverts par des approches sans *a priori*, ont conduit à l'adoption de la métagénomique par une partie des pathologistes au début du XXI^e siècle. L'idée générale a été d'appliquer les

approches de métagénomique à la pathologie en déplaçant le cadre d'étude de l'écosystème à l'organisme. Le terme de métagénomique est donc ici légèrement déformé par rapport à son acception première, et certains scientifiques préfèrent utiliser le terme de « nouvelles techniques de séquençage » (NextGEN ou NGS, *next generation sequencing*) pour nommer cette nouvelle approche. Nous garderons ici le terme de métagénomique appliquée à la pathologie car l'utilisation des nouvelles techniques de séquençage se trouve ici à nouveau couplée à des méthodes d'extraction et d'amplification sans *a priori*. Ces travaux ont particulièrement fleuri en virologie, ce qu'illustre la croissance exponentielle des publications depuis 2005 (Figure 1). Nous nous concentrerons donc dans cette revue sur les principaux résultats obtenus depuis 2001 et les travaux pionniers de T. Allander *et al.* [7] en métagénomique virale.

Les premières études basées sur la métagénomique virale ont mis l'accent sur la découverte de nouveaux virus [10]. Mais depuis, plusieurs avancées ont été réalisées dans d'autres disciplines de la virologie grâce aux résultats issus de travaux de métagénomique virale et, plus largement, de l'utilisation des NGS. Nous discuterons ici de ces avancées dans quatre champs : la découverte de nouveaux virus, le diagnostic, l'épidémiologie moléculaire et l'évolution (Figure 2).

Métagénomique virale et découverte de nouveaux virus

Dès 2008, un papier *princeps* a illustré le potentiel des approches métagénomiques de criblage à haut débit (*high-throughput screening*, HTS) en identifiant le

Figure 2. Principales étapes des travaux de métagénomique. Les techniques de séquençage à haut débit (454 [Roche] et Illumina [Solexa]) sont de nos jours majoritairement utilisées dans ces travaux. Pour la technique 454 (Roche), six étapes ont lieu successivement. **A, B.** Construction d'une librairie avec dénaturation de l'ADN et fixation d'adaptateurs ; **C.** Fixation de l'ADN sur des microbilles contenant à leur surface des oligonucléotides complémentaires aux adaptateurs ; **D.** Emulsion PCR ; **E.** Pyroséquençage dans des plaques en fibre optique picotitrée. Chaque nucléotide incorporé dans la synthèse du brin complémentaire libère un groupe pyrophosphate qui engendre la production d'un signal lumineux. **F.** Lecture des signaux lumineux aboutissant à la production d'un pyrogramme indiquant la séquence obtenue. La technique Illumina (Solexa) se déroule en cinq étapes successives. **G.** Dénaturation de l'ADN, fixation d'adaptateurs, puis hybridation des fragments d'ADN avec les oligonucléotides complémentaires des adaptateurs sur un support ; **H.** Les brins se courbent et s'hybrident avec les oligonucléotides proches. Le brin complémentaire est synthétisé par PCR. **I.** Les ADN double-brins sont dénaturés et les processus **G** et **H** reprennent ; **J-K.** L'incorporation de nucléotides marqués par des fluorochromes durant la synthèse d'ADN permet une lecture de la séquence. Les séquences obtenues à haut débit sont identifiées par similarité de leurs séquences (Blast, *basic local alignment search tool*) avec celles d'une base de données internationale (GenBank, NCBI). Les résultats de ces identifications de séquences sont exploitables au sein de quatre disciplines de la virologie : (1) la description de nouveaux virus qui implique des travaux de systématique et de taxonomie, (2) l'épidémiologie moléculaire avec pour corollaire la surveillance virale, (3) le diagnostic viral et (4) l'évolution virale.

GLOSSAIRE

Contig : séquence obtenue à la suite de l'assemblage de séquences contiguës présentant des zones identiques et chevauchantes à leurs extrémités 3' et 5'.

Écogénomique : méthode qui consiste à « étiqueter » les séquences de chaque échantillon avant le séquençage de manière à pouvoir réattribuer chaque séquence à son échantillon d'origine à la suite du séquençage à haut débit.

Métagénomique : analyse sans étape de culture de l'ensemble des acides nucléiques des microorganismes présents dans un milieu donné. Le catalogue de ces séquences représente le génome collectif ou le génome global d'un échantillon donné, appelé métagénome.

Microbiome : ensemble des génomes des bactéries colonisant l'organisme d'un animal.

NGS-HTS (next generation sequencing/high throughput sequencing) : nouvelles techniques de séquençage permettant la production d'un grand nombre de séquences rapidement et à moindre coût par comparaison avec la méthode Sanger (méthodes post-Sanger). Elles permettent notamment le séquençage d'un mélange de séquences présentes dans un échantillon.

Pipeline : procédé qui permet de lier et automatiser de façon indépendante les différentes étapes de traitement d'une instruction par le processeur comme par exemple l'ensemble des traitements bio-informatiques successifs visant à analyser les jeux de données issus du séquençage à haut débit.

Reads : lectures de fragments nucléotidiques. Par exemple, un cycle de 454 FLX délivre environ un million de lectures de 700 pb chacune en 20 h, soit 900 Mb de données.

Virome : ensemble des génomes d'une population virale trouvés dans un même organisme ou dans un même environnement.

premier polyomavirus oncogène humain, le virus de Merkel [17]. Dans la tumeur de Merkel (une tumeur d'une cellule du système neuroendocrine à la base de l'épiderme), des transcrits codant pour une protéine similaire à l'oncogène T du virus SV40 ont été identifiés [41]. Le virus découvert sur cette base est présent à la surface de la peau chez 70 % des individus, mais, chaque année, seuls un à trois individus sur un million développeront un cancer. Un des mérites de ce travail est de démontrer que des virus extrêmement fréquents peuvent rester inconnus longtemps et être révélés par des approches de ce type. À ce jour, c'est le seul nouveau virus oncogène détecté par ce type d'approche et, *a posteriori*, on ne peut que s'étonner qu'il ait été aussi le premier virus humain découvert par ce type de technologie.

La découverte de virus inconnus est rendue difficile par le fait qu'il n'existe pas de séquences universellement conservées chez les virus. Cet écueil méthodologique a probablement eu pour effet de stimuler les approches indépendantes de séquences en pathologie virale. Pour découvrir des virus inconnus, il faut par ailleurs être capable d'assembler *de novo* de longs « contigs » (blocs de séquences continues, voir *Glossaire*) au sein d'échantillons biologiques complexes. En effet, seuls des contigs de taille relativement grande peuvent permettre d'identifier comme significatifs des pourcentages d'identité très faibles avec des virus connus, voire

d'identifier des motifs génomiques indépendants des homologies de séquence. On peut néanmoins prédire que le nombre de virus distants découverts va augmenter avec le développement récent de séquenceurs conjuguant haut débit et longs *reads* (lectures de fragments de séquence nucléotidique, voir *Glossaire*), de manière à pouvoir identifier des virus distants malgré leurs titres faibles.

Pour l'instant, c'est dans son application aux infections virales productives que ce type d'approche s'est révélé le plus performant. Par exemple, l'étude du virome fécal de nombreuses espèces a permis d'identifier un nombre croissant de virus nus, particulièrement au sein de la famille des *Circoviridae*. La résistance de ces virus et leur excrétion avec un fort titre dans un matériel biologique qu'il est possible de concentrer relativement facilement sont pour beaucoup dans cette accumulation de connaissances [18]. Nous avons ainsi identifié une nouvelle espèce virale dans la famille des *Picornaviridae*, qui définit un nouveau genre viral que nous avons appelé *Pasivirus* (*Parecho Sister Clade*), compte tenu de sa proximité avec les parechovirus [19]. De même, à la surface de la peau humaine, notre équipe a identifié différents virus au sein des *Polyomaviridae* (HPyV9) [20] et *Circoviridae* (genre *Gyrovirus*) [21]. La flore virale des poumons [22] et celle de la peau [23] se sont également avérées très riches (*Figure 3*). Dans la plupart de ces études qui concernaient des syndromes où ces virus ont été recherchés *a posteriori*, aucune relation ultérieure de ce virus avec une maladie n'a pu être mise en évidence.

Au-delà du cas des infections inapparentes, et hormis le cas remarquable du virus de Merkel évoqué plus haut, l'identification de virus responsables de maladies s'est révélée beaucoup plus facile dans les maladies aiguës. C'est le cas de l'identification du virus de Schmallenberg, un nouvel orthobunyavirus émergeant en Europe chez les ruminants [24], du picornavirus de l'hépatite de la dinde [25] et de l'arterivirus hémorragique simien [26]. À l'inverse et jusqu'à présent, le virus de Merkel reste le seul virus identifié par NGS responsable de maladies chroniques incluant les cancers. De manière générale, les implications biologiques et thérapeutiques de la découverte fortuite ou orientée de séquences virales dans un syndrome infectieux nécessiteront un travail d'imputabilité, qui doit reposer sur une actualisation des postulats de Koch¹ que certains ont pu proposer [27]. Cette imputabilité pourrait être difficile à établir pour des infections persistantes n'évoluant vers des conséquences pathologiques que chez un petit nombre d'individus.

¹ Postulats de Koch : (1) l'agent doit être présent chez tous les individus atteints ; (2) l'agent peut être cultivé ; (3) l'agent introduit chez l'hôte, ou dans un modèle animal proche, induit la maladie ; (4) l'agent peut être à nouveau isolé à partir d'un animal présentant la pathologie.

Figure 3. Analyse métagénomique de la flore virale au niveau de la peau. L'abondance relative des lectures (reads) de papillomavirus (β-HPV, γ-HPV et HPV inconnus), de circovirus et d'autres virus à ADN simple brin, ainsi que des polyomavirus (MCPyV, HPyV6 et 7), a été mesurée sur la peau de six patients dont cinq patients sains et un patient malade présentant une tumeur de Merkel due au virus de Merkel (MCPyV) (patient 100066) [23]. ssDNA : single-stranded DNA.

Application de la métagénomique virale au diagnostic

L'utilisation des NGS dans une approche diagnostique apparaissait jusqu'à il y a peu comme totalement irréaliste en raison de la combinaison de coûts élevés et d'une durée longue de séquençage et d'analyse. Cette perception est en train de changer pour plusieurs raisons.

L'apparition de séquenceurs de paillasse (Illumina MiSeq, Ion Torrent PGM) réduit la durée de séquençage à moins de 24 h au prix néanmoins d'une diminution du débit sans doute critique pour un certain nombre d'indications. Encore plus récemment, le Proton de Ion Torrent a été mis sur le marché : il est doté d'un

Figure 4. Analyse d'un écouvillonnage nasal par la technologie Illumina lors d'une pneumopathie humaine. Deux virus ont été mise en évidence : le virus respiratoire syncytial (RSV) et le cytomégalovirus (CMV) dont les génomes quasi complets ont été acquis. L'assemblage *de novo* (assembleur CLC Genomics) a identifié des contigs homologues chez ces deux virus. L'ensemble des *reads* non assemblés ont alors été localisés sur ces génomes de référence (mappeur CLC). Chaque graphe présente en abscisse les positions nucléotidiques du génome et en ordonnée la couverture du séquençage, c'est-à-dire le nombre de fois où chaque position est séquencée (source Pathoquest et Institut Pasteur du Cameroun [P. Buchy]).

tage d'être (pratiquement) sans *a priori* mais, malheureusement, elles ne détectaient que les micro-organismes cultivables. La métagénomique pérennise l'avantage d'une détection large sans *a priori*, mais sans la conditionner à des capacités d'isolement.

Une crainte récurrente est que ce type d'approche ne trouve que des agents infectieux sans impact médical, et soit au final plus une source

débit considérable pour une durée de séquençage de l'ordre de quatre heures. Avec une bonne optimisation des outils de bio-informatique, obtenir une réponse en deux jours deviendra alors possible, pour un coût acceptable dans le cadre d'une utilisation hospitalière réservée à certains malades.

La sensibilité de certains *pipelines* (ce terme désignant l'ensemble des séquences allant du traitement de l'échantillon jusqu'à l'analyse bio-informatique des séquences, voir *Glossaire*) est au moins équivalente à celle de la PCR (*polymerase chain reaction*) ciblée, mais sans qu'il soit nécessaire de définir *a priori* des cibles. Nous avons utilisé récemment ce *pipeline* dans différents syndromes infectieux chez l'homme (encéphalites, pneumopathies), préalablement criblés négativement avec les outils classiques de PCR. À nouveau, le diagnostic par métagénomique nécessite d'assembler *de novo* de longs contigs au sein d'échantillons biologiques complexes (Figure 4). La métagénomique met en évidence des agents pathogènes connus mais qui n'ont pas été identifiés par PCR, souvent en raison d'amorces ne couvrant pas l'étendue des séquences possibles. Elle représentera sans nul doute rapidement une alternative « moderne » aux techniques de culture maintenant abandonnées. Celles-ci avaient l'immense avan-

de confusion qu'une aide médicale. Tout d'abord, il est aisé de filtrer les résultats de manière à ce que ne soit délivrée au médecin que la liste des agents pathogènes présents. Ensuite, si l'on adopte une vision prospective, certaines associations de virus, bactéries et champignons révélées par la métagénomique devraient avoir une valeur diagnostique et orienter de manière significative le traitement. Il est certain que l'utilisation de ces outils non biaisés, communs à tous les types de microorganismes, va ouvrir un nouveau champ dans le diagnostic des maladies infectieuses.

Métagénomique virale et connaissance épidémiologique

Dans un autre champ, la métagénomique virale est en train de bousculer la connaissance épidémiologique des infections d'intérêt sanitaire, notamment en redéployant les territoires géographiques d'échantillonnage et en prenant mieux en compte les zones sauvages [28]. Le nombre

d'espèces séquencées par cette approche augmente en effet considérablement la quantité d'informations génétiques disponible. Les reconstructions phylogénétiques qui sont déduites de ces informations peuvent inférer avec davantage de fiabilité les liens épidémiologiques entre les isolats à l'échelle spatiale et chronologique [29], et permettre d'établir des chaînes de transmission inter-hôtes [30]. Certaines études ont permis d'appréhender des événements liés à un franchissement d'espèces et au développement d'une émergence. À titre d'exemple, une étude de séquençage profond des virus *Influenza* de type A a montré que le passage du réservoir canard vers des hôtes aviaires inhabituels (poulet, dinde) s'accompagne souvent de l'émergence de modifications génétiques des virus [31]. Dans cette étude, les auteurs montrent l'émergence de deux types de délétions dans le gène codant pour la neuraminidase qui confèrent probablement aux nouveaux génotypes viraux un avantage adaptatif chez des hôtes aviaires inhabituels, notamment chez la dinde [31]. Par ailleurs, depuis 2009, une nouvelle génération de travaux a émergé dans le domaine de la pathologie des plantes ; ces données permettent non seulement d'analyser le génome global d'un écosystème ou d'un organisme, mais aussi de relier directement les séquences des agents pathogènes à leur hôte et/ou à une position géographique [32]. Ces travaux *princeps* d'écogénomique ont révolutionné la vision de la distribution des phytovirus en révélant que près de 70 % des plantes analysées étaient « virosées » (infectées par un virus) [32]. Cette innovation récente, couplée au fait que les plantes sont immobiles et donc « ré-échantillonnables » dans le temps, devrait permettre de mieux comprendre la dynamique spatiotemporelle de la diversité phytovirale d'un agro-écosystème. Enfin, grâce à l'utilisation d'algorithmes probabilistes de plus en plus performants, les données générées par NGS constitueront une source d'information inestimable pour modéliser la dynamique des infections virales et prédire les émergences. Ces études devraient permettre de mieux rationaliser les approches plus coûteuses de collecte de données épidémiologiques sur le terrain.

Métagénomique virale et dynamique d'évolution des pathogènes

Les virus, en particulier ceux qui ont un intérêt médical, se caractérisent généralement par un nombre élevé d'individus par génération, des cycles de réplication très courts et un fort taux de mutations. L'accumulation des diversités qui en résulte, observable à l'échelle d'une vie humaine, fait des virus des modèles de choix pour étudier l'évolution du vivant. L'évolution des virus a par ailleurs un impact en virologie médicale, avec des conséquences sur la pathogénicité, le franchissement de la barrière d'espèce, l'échappement aux antiviraux et aux vaccins. Le niveau de diversité généré dépend du type de virus concerné et des mécanismes en jeu. Ainsi, les virus à ARN simple brin sont les plus « plastiques » : le taux de substitution/site/an est de 10^{-3} à 10^{-5} , soit 4 à 6 $\log_{10}$ au-dessus du taux de substitution généralement observé chez les organismes pluricellulaires [33]. Par ailleurs, les recombinaisons ou les échanges de segments participent à la génération de cette diversité. Pour appréhender la dynamique d'évolution des pathogènes et la dispersion des populations virales (quasi-espèces), les premières études reposaient sur des approches lourdes de clonage suivi du séquençage des clones par la technique de Sanger. Les NGS ou HTS

permettent désormais d'explorer directement la diversité virale *in situ*, pour peu que certaines précautions d'ordre méthodologique soient prises pour éliminer les erreurs d'amplification, de séquençage et surtout d'assemblage des séquences qui peut produire des chimères artificielles [34]. Dans les domaines humains mais aussi vétérinaires, les NGS se développent pour étudier la diversité virale générée aussi bien par substitution que par recombinaison [34]. Des travaux sur les cardiovirus humains ont ainsi mis en évidence la trace d'une recombinaison avec un theilovirus de rat, suggérant que l'ancêtre commun de ces virus proviendrait du rat [35]. La puissance de ces méthodes permet par ailleurs de caractériser la dynamique évolutive des virus chez un hôte ou entre différents hôtes [36]. La génération de la diversité peut également être explorée dans différents sites chez un même hôte et ce en fonction du développement de l'infection [37]. Les NGS permettent aussi de caractériser de façon dynamique l'émergence de mutants échappant aux antiviraux ou aux vaccins [38]. Au-delà de la connaissance qu'elles révèlent sur l'évolution d'un virus en particulier, les NGS contribueront certainement, dans un proche avenir, à élucider certaines étiologies complexes chez l'homme, l'animal ou la plante, pouvant résulter d'interactions subtiles (recombinaison, complémentation, etc.) entre variants d'une même espèce virale ou entre virus co-infectant une même cellule hôte.

Conclusions

Les travaux de métagénomique virale, ou plus largement le mouvement actuel d'acquisition massive de jeux de données NGS, sont confrontés à un double défi : premièrement, le traitement bio-informatique de ces millions voire milliards de séquences (tri, inventaire, stockage, accessibilité, etc.) ; deuxièmement, les interprétations biologique, écologique ou thérapeutique de ces nouveaux jeux de données. On peut cependant entrevoir que les progrès qui seront réalisés dans l'organisation et l'intégration de cette masse de données permettront de dépasser un niveau purement descriptif des résultats en leur donnant du sens dans le contexte des différents fronts de recherche et selon une démarche scientifique. Les données qualitatives actuellement obtenues seront enrichies par des données quantitatives robustes basées sur les répliquations d'analyses dans le temps et dans l'espace, et soumises à des analyses statistiques rigoureuses. Nonobstant ces obstacles, la métagénomique a permis, au cours de cette dernière décennie, de réévaluer plusieurs concepts en écologie et en pathologie. Le rôle des virus dans un organisme ou dans un écosystème est par exemple un sujet à nouveau débattu, avec l'idée que des virus mutualistes ou bien

commensaux existent probablement [39]. Cette nouvelle typologie du rôle des virus, associée aux nouveaux ordres de grandeur de la diversité virale observée chez un hôte, nécessitera de mieux prendre en compte les interactions entre virus présents au sein des hôtes eucaryotes et, pour les virus pathogènes, d'actualiser les postulats de Koch [27]. La pathologie, dont l'étude s'est enrichie au début du XXI^e siècle d'une approche développée initialement en écologie, devra sûrement accentuer son rapprochement avec cette discipline scientifique afin de mieux comprendre l'évolution et la dynamique des virus pathogènes ou mutualistes qui colonisent les humains, les animaux et les plantes. ♦

SUMMARY

Pathology and viral metagenomics, a recent history

Human, animal and plant viral diseases have greatly benefited from recent metagenomics developments. Viral metagenomics is a culture-independent approach used to investigate the complete viral genetic populations of a sample. During the last decade, metagenomics concepts and techniques that were first used by ecologists progressively spread into the scientific field of viral pathology. The sample, which was first for ecologists a fraction of ecosystem, became for pathologists an organism that hosts millions of microbes and viruses. This new approach, providing without *a priori* high resolution qualitative and quantitative data on the viral diversity, is now revolutionizing the way pathologists decipher viral diseases. This review describes the very last improvements of the high throughput next generation sequencing methods and discusses the applications of viral metagenomics in viral pathology, including discovery of novel viruses, viral surveillance and diagnostic, large-scale molecular epidemiology, and viral evolution. ♦

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

- Handelsman J, Rondon MR, Brady SF, et al. Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products. *Chem Biol* 1998 ; 5 : R245-9.
- Schmidt TM, Delong EF, Pace NR. Analysis of a marine picoplankton community by 16S ribosomal-RNA gene cloning and sequencing. *J Bacteriol* 1991 ; 173 : 4371-8.
- Pace NR, Stahl DA, Lane DJ, et al. Analyzing natural microbial populations by rRNA sequences. *ASM News* 1985 ; 51 : 4-12.
- Pace NR. A molecular view of microbial diversity and the biosphere. *Science* 1997 ; 276 : 734-40.
- Amann RL, Ludwig W, Schleifer KH. Phylogenetic identification and *in-situ* detection of individual microbial-cells without cultivation. *Microbiol Rev* 1995 ; 59 : 143-69.
- Venter JC, Remington K, Heidelberg JF, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 2004 ; 304 : 66-74.
- Allander T, Emerson SU, Engle RE, et al. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci USA* 2001 ; 98 : 11609-14.
- Breitbart M, Salamon P, Andresen B, et al. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci USA* 2002 ; 99 : 14250-5.
- Rosario K, Breitbart M. Exploring the viral world through metagenomics. *Curr Opin Virol* 2011 ; 1 : 1-9.
- Djikeng A, Kuzmickas R, Anderson NG, et al. Metagenomic analysis of RNA viruses in a fresh water lake. *PLoS One* 2009 ; 4 : e38499.
- DeLong EF, Preston CM, Mincer T, et al. Community genomics among stratified microbial assemblages in the ocean's interior. *Science* 2006 ; 311 : 496-503.
- Dupré J, O'Malley MA. Metagenomics and biological ontology. *Stud Hist Philos Biol Biomed Sci* 2007 ; 38 : 834-46.
- Methe BA, Nelson KE, Pop M, et al. A framework for human microbiome research. *Nature* 2012 ; 486 : 215-21.

- Faust K, Sathirapongsasuti JF, Izard J, et al. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol* 2012 ; 8 : e1002606.
- Dupré J. Emerging sciences and new conceptions of disease; or, beyond the monogenomic differentiated cell lineage. *Euro J Phil Sci* 20121 ; 1 : 119-31.
- Lai YP, Di Nardo A, Nakatsuji T, et al. Commensal bacteria regulate Toll-like receptor 3-dependent inflammation after skin injury. *Nat Med* 2009 ; 15 : 1377-U4.
- Feng HC, Shuda M, Chang Y, et al. Clonal integration of a polyomavirus in human Merkel cell carcinoma. *Science* 2008 ; 319 : 1096-100.
- Phan TG, Kapusinszky B, Wang CL, et al. The fecal viral flora of wild rodents. *PLoS Pathog* 2011 ; 7 : e1002218.
- Sauvage V, Le Gouil M, Cheval J, et al. A member of a new Picornaviridae genus is shed in pig feces. *Int J Infect Diseases* 2012 ; 16 : E456.
- Sauvage V, Foulongne V, Cheval J, et al. Human polyomavirus related to African green monkey lymphotropic polyomavirus. *Emerg Infect Dis* 2011 ; 17 : 1364-70.
- Sauvage V, Cheval J, Foulongne V, et al. Identification of the first human Gyrovirus, a virus related to chicken anemia virus. *J Virol* 2011 ; 85 : 7948-50.
- Willner D, Furlan M, Haynes M, et al. Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PLoS One* 2009 ; 4 : e7370.
- Foulongne V, Sauvage V, Hebert C, et al. Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PLoS One* 2012 ; 7 : e38499.
- Hoffmann B, Scheuch M, Hoper D, et al. Novel Orthobunyavirus in cattle, Europe, 2011. *Emerg Infect Dis* 2012 ; 18 : 469-72.
- Honkavuori KS, Shivaprasad HL, Briese T, et al. Novel Picornavirus in turkey poult with hepatitis, California, USA. *Emerg Infect Dis* 2011 ; 17 : 480-7.
- Lauck M, Hyeroba D, Tumukunde A, et al. Novel, divergent simian hemorrhagic fever viruses in a wild Ugandan red colobus monkey discovered using direct pyrosequencing. *PLoS One* 2011 ; 6 : e19056.
- Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* 2012 ; 2 : 63-7.
- Van den Brand JMA, van Leeuwen M, Schapendonk CM, et al. Metagenomic analysis of the viral flora of pine marten and European badger feces. *J Virol* 2012 ; 86 : 2360-5.
- Pybus OG, Suchard MA, Lemey P, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci USA* 2012 ; 109 : 15066-71.
- Escobar-Gutierrez A, Vazquez-Pichardo M, Cruz-Rivera M, et al. Identification of hepatitis C virus transmission using a Next-Generation Sequencing approach. *J Clin Microbiol* 2012 ; 50 : 1461-3.
- Croville G, Soubies SM, Barbieri J, et al. Field monitoring of avian influenza viruses: whole-genome sequencing and tracking of neuraminidase evolution using 454 pyrosequencing. *J Clin Microbiol* 2012 ; 50 : 2881-7.
- Roossinck MJ, Saha P, Wiley G, et al. Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol* 2010 ; 19 : 81-8.
- Duffy S, Shackleton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 2008 ; 9 : 267-76.
- Beerenwinkel N, Günthard HF, Roth V, et al. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. *Front Microbiol* 2012 ; 3 : 1-16.
- Blinkova O, Kapoor A, Victoria J, et al. Cardioviruses are genetically diverse and cause common enteric infections in South Asian children. *J Virol* 2009 ; 83 : 4631-41.
- Bull RA, Eden JS, Luciani F, et al. Contribution of intra- and interhost dynamics to Norovirus evolution. *J Virol* 2012 ; 86 : 3219-29.
- Wright CF, Morelli MJ, Thebaud G, et al. Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. *J Virol* 2011 ; 85 : 2266-75.
- Nishijima N, Marusawa H, Ueda Y, et al. Dynamics of hepatitis B virus quasispecies in association with nucleos(t)ide analogue treatment determined by ultra-deep sequencing. *PLoS One* 2012 ; 7 : e35052.
- Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol* 2011 ; 9 : 99-108.
- Jordan B. Une révolution longuement attendue. *Med Sci (Paris)* 2008 ; 24 : 869-73.
- Kanitakis J. Carcinome à cellules de Merkel : première implication connue d'un polyomavirus dans la genèse d'un cancer humain. *Med Sci (Paris)* 2008 ; 24 : 570-1.

TIRÉS À PART
P. Roumagnac