Éditorial

Virologie

Virologie 2014, 18 (5) : 247-50

Diagnostic des viroses et séquençage haut débit : vers un changement de paradigme ?

Sébastien Massart[1]
Olmos Antonio[2]
Haissam Jijakli[1]
Thierry Candresse[3,4]

[1] Laboratoire de phytopathologie, Université de Liège, Gembloux Agro-Bio Tech, passage des Déportés, 2, 5030 Gembloux, Belgium <sebastien.massart@ulg.ac.be>
[2] Centro de Protección Vegetal, Instituto Valenciano de Investigaciones Agrarias (IVIA), Apartado Oficial, 46113 Moncada, Valencia, Spain
[3] UMR 1332 de Biologie du fruit et Pathologie, INRA, CS20032, 33882 Villenave d'Ornon cedex, France
[4] UMR 1332 de Biologie du fruit et Pathologie, Université de Bordeaux, CS20032, 33882 Villenave d'Ornon cedex, France

Pour être adoptée en routine, une technique de diagnostic de virus doit idéalement présenter les caractéristiques suivantes : techniquement robuste, reproductible et répétable, diagnostic sensible et spécifique, facilité de mise en œuvre, y compris pour des utilisateurs peu formés ou inexpérimentés, automatisable, rapide, économiquement compatible avec les besoins de la filière et basée sur des standards industriels généralisés [1]. L'Elisa (*Enzyme Linked ImmunoSorbent Assay*), apparue dans les années 1970, et la PCR en temps réel, apparue à la fin des années 1990, possèdent la plupart de ces caractéristiques et sont ainsi devenues les deux standards utilisés de manière intensive pour le diagnostic des viroses, en particulier dans le domaine végétal. Pour chacune de ces techniques, une dizaine d'années ont été nécessaires entre les premiers développements et leur utilisation par les laboratoires de diagnostic de routine. Ces technologiques ont ainsi amélioré de manière significative notre capacité à diagnostiquer des infections virales. Cependant, leur utilité est limitée par la nécessité de déterminer a priori les virus ciblés par le test et par la disponibilité d'anticorps ou de séquences nucléotidiques spécifiques. Elles ne permettent donc pas le diagnostic de virus inconnus ou de souches virales divergentes peu ou mal caractérisées. L'indexage complet, c'est-à-dire l'identification de tous les virus présents dans un échantillon, était donc un objectif inatteignable jusqu'à tout récemment.

Les technologies de séquençage haut débit se sont développées de manière exponentielle ces dernières années, le prix au nucléotide lu diminuant de moitié tous les semestres. Leur démocratisation ouvre de nouvelles perspectives de recherche et d'application dans de nombreux domaines de la virologie, dont le diagnostic. Plusieurs plateformes technologiques pour le séquençage haut débit existent à l'heure actuelle, d'autres apparaissent régulièrement. Elles sont décrites plus en détail par Barba, Czosnek [2]. Toutes ces technologies reposent sur 3 étapes clefs : la préparation des échantillons en vue du séquençage (création des librairies), la multiplication clonale des librairies préparées et le séquençage massif des librairies amplifiées. L'écueil principal du séquençage haut débit n'est plus la génération des données de séquence elle-même mais leur traitement et leur analyse par bioinformatique. Les questions informatiques qui se posent sont nouvelles par rapport aux techniques de diagnostic actuelles. En effet, le diagnostic d'un agent viral par Elisa ou PCR en temps réel repose sur l'observation d'une coloration ou d'une émission de fluorescence. Le traitement de l'image et l'interprétation des résultats sont des étapes simples et immédiates. Ce n'est pas le cas pour les données de séquençage haut débit qui doivent passer par plusieurs étapes de traitement pouvant éventuellement introduire des artéfacts. L'analyse bioinformatique peut-être schématiquement subdivisée en 4 étapes : le contrôle qualité des séquences obtenues, leur assemblage en contigs, la caractérisation des contigs et l'identification de variations chez ou entre les échantillons. Il est important de souligner que ces analyses bioinformatiques deviennent elles aussi techniquement et économiquement plus abordables grâce aux progrès considérables des logiciels d'analyse commerciaux ou en libre accès.

doi:10.1684/vir.2014.0576

Pour citer cet article : Massart S, Antonio O, Jijakli H, Candresse T. Diagnostic des viroses et séquençage haut débit : vers un changement de paradigme ? *Virologie* 2014; 18(5) : 247-50 doi:10.1684/vir.2014.0576

Grâce à l'ensemble de ces développements, il est maintenant conceptuellement possible de réaliser l'indexage complet d'un échantillon en vue d'identifier toute séquence virale présente, indépendamment de son origine. Les techniques de séquençage haut débit sont ainsi déjà couramment utilisées dans les laboratoires de recherche en vue de tenter d'identifier l'agent causal de maladies ou syndromes d'origine encore inconnue et ont par exemple permis l'identification de plusieurs dizaines de nouveaux virus, notamment de plante, ces dernières années [2]. Au-delà de ces avancées dans notre compréhension de l'étiologie des maladies, le séquençage haut débit est également susceptible de modifier profondément la manière de réaliser et de considérer le diagnostic des viroses, végétales ou autres, dans les prochaines années. La transition du séquençage haut débit vers une méthodologie de diagnostic de routine soulève de nombreuses questions techniques, bioinformatiques, scientifiques et réglementaires. Ces questions se posent tant pour les viroses végétales [3] que pour le diagnostic clinique des virus humains [4]. Elles peuvent donc être considérées comme intrinsèques à ces technologies. Nous soulignerons plus particulièrement dans ce qui suit les spécificités liées au diagnostic phytopathologique.

Une première question technique est celle de la stratégie de séquençage utilisée, et plus particulièrement de la nature des acides nucléiques séquencés. Différentes approches ont été décrites à ce jour, chacune avec ses avantages et ses inconvénients. Les facteurs clefs dans le choix d'un futur protocole de diagnostic sont la facilité de préparation des échantillons, la possibilité d'utiliser des méthodes standardisées dans l'industrie du séquençage et la possibilité d'identifier un spectre de virus le plus large possible. Aucune des méthodes actuelles ne semble réunir toutes ces qualités, même si celles du séquençage des petits ARN interférents (siRNA, 21-24 nt) et des ARN messagers sont actuellement les plus travaillées [3]. Elles ont l'avantage de pouvoir détecter un large spectre de virus indépendamment de leur nature (ADN ou ARN) et de leur structure (linéaire, circulaire, simple ou double brin). Elles reposent sur des protocoles standards, simples et automatisables de préparation de librairies mais nécessitent la génération d'une grande quantité de séquences à cause de la dilution des séquences virales dans un très grand nombre de séquences de l'hôte. Chez les plantes, l'absence de fluide acellulaires, qui permettrait l'extraction de populations d'acides nucléiques pauvres en séquences de l'hôte (comme le plasma ou le fluide cérébrospinal chez les animaux ou l'homme), rend ce problème aigu en diagnostic phytopathologique. Néanmoins, la diminution continue des coûts de séquençage et le progrès des techniques bioinformatiques minimiseront cet inconvénient dans le futur. Indépendamment de cette évolution, une comparaison directe des approches ciblant différentes populations d'acides nucléiques sera nécessaire pour déterminer précisément la technique la plus appropriée à chaque questionnement.

La sensibilité d'une méthode de diagnostic est cruciale dans certaines applications comme la production de plants ou de semences certifiés indemnes de virus. Pour les virus humains, une sensibilité similaire à la PCR en temps réel peut être obtenue mais elle dépend de l'obtention d'un nombre suffisant de séquences [5]. Des résultats prometteurs ont par ailleurs été obtenus pour la détection d'un virus de plante, le *Tomato spotted wilt virus* (TSWV) [6]. Des essais comparatifs sur des échantillons artificiellement contaminés par un ou plusieurs virus devraient permettre de déterminer la quantité minimale de séquences à générer (profondeur de séquençage) pour égaler ou surpasser la sensibilité de la PCR en temps réel. D'après des données récentes, il sera probablement nécessaire d'ajuster cette profondeur de séquençage en fonction du matériel analysé. En effet, la fraction de siRNA d'origine virale semble être beaucoup plus faible chez les espèces ligneuses que chez les espèces herbacées. De plus, certaines plantes produisent des composés qui interfèrent avec l'extraction et l'efficacité des enzymes utilisées lors de la préparation des librairies et nécessiteront donc le développement de protocoles spécifiques.

La problématique de fixation d'un seuil de détection est un autre facteur-clef de toute technique de diagnostic mais elle prend une dimension particulière dans le séquençage à haut débit. En effet, certains virus présents à très faible concentration pourraient n'être identifiés que par la présence de quelques séquences, voire d'une unique séquence sur plusieurs millions de séquences analysées. Ce type de résultat pourrait également provenir d'une contamination d'origine environnementale ou accidentelle au sein du laboratoire de diagnostic. La problématique des contaminations et l'importance d'une bonne organisation des laboratoires seront donc tout aussi cruciales que pour la PCR en temps réel. Certaines stratégies sont déjà envisagées pour détecter de possibles contaminations en tenant compte de la couverture du génome de l'espèce détectée. Ainsi, une faible couverture des séquences mais sur l'ensemble du génome signerait probablement la présence d'un agent viral à faible concentration tandis qu'une couverture plus importante mais sur une seule ou quelques régions génomiques serait un indice de contamination.

L'utilisation du séquençage haut débit comme technique de diagnostic nécessitera par ailleurs la disponibilité d'outils bio-informatiques simples, puissants et faciles à utiliser. Les programmes actuels peuvent notamment provoquer l'identification erronée de recombinants entre familles de virus suite à des erreurs d'analyse des séquences [7]. Les progrès réalisés dans ce domaine ces dernières années sont néanmoins remarquables mais de nombreux efforts doivent encore être faits pour le développement de programmes

plus rapides, plus conviviaux et permettant l'identification de toute séquence virale, même très divergente des virus connus. La fiabilité de ces programmes devra de plus être compatible avec les exigences du diagnostic.

La répétabilité, la reproductibilité et la robustesse des protocoles de séquençage à haut débit, incluant les étapes de génération des séquences au laboratoire et leur analyse bioinformatique, n'ont pas encore été évaluées par des validations intra- et inter-laboratoires. Une telle démarche de validation inter-laboratoires, permettant notamment de connaître les rapports de vraisemblance (*Likelihood ratios*) critiques pour une évaluation objective de la fiabilité des essais [8], est pourtant devenue indispensable avant de proposer une nouvelle technique de référence pour le diagnostic.

Bien que beaucoup plus complexes, ces questions techniques sont similaires à celles qui se sont posées lors du développement de l'Elisa ou de la PCR comme techniques de diagnostic. Par contre, l'avènement du séquençage à haut débit comme technique d'identification de virus et son utilisation potentielle comme outil de diagnostic posent également des questions scientifiques totalement originales. Le rythme de découverte de nouveaux virus s'est considérablement accéléré ces dernières années. De nouveaux virus sont ainsi découverts et décrits chaque mois, notamment à partir de tissus asymptomatiques de plantes cultivées ou de plantes sauvages. L'impact de ces découvertes nécessite une réflexion en profondeur sur la manière de concevoir la taxonomie virale et sur la caractérisation biologique des agents nouvellement identifiés. La taxonomie virale sera très probablement fortement impactée par le séquençage à haut débit, les virus restant à découvrir étant sans nul doute bien plus nombreux que ceux qui sont actuellement connus. Ainsi, le Comité international sur la taxonomie des virus (ICTV) a récemment accepté le principe d'inclure de nouvelles espèces virales uniquement sur la base d'une séquence génomique complète [9], en l'absence de tous autres éléments de caractérisation biologique. Par ailleurs, lors de la détection de la séquence complète ou partielle d'une espèce virale nouvelle ou non caractérisée, le diagnosticien sera confronté à une interprétation et, éventuellement, à des décisions difficiles. En effet, certains virus peuvent être latents ou commensaux, tant dans le règne végétal qu'animal, et certains pourraient même avoir des effets bénéfiques sur le phénotype de l'hôte. Dans un tel contexte, comment interpréter la détection d'un nouvel agent pour lequel on ne dispose d'aucune information biologique quant à son impact ? Par ailleurs, la détection de nouvelles séquences virales dans un échantillon peut ne pas refléter nécessairement l'infection de l'hôte mais, éventuellement, l'infection d'organismes associés à cet hôte. L'infection par des mycovirus de champignons endophytes de plantes en représente un exemple banal [3]. Ces observations renforcent la nécessité d'une caractérisation approfondie des propriétés biologiques et de l'écologie des virus nouvellement identifiés, y compris dans les espèces végétales naturellement présentes dans les écosystèmes. Cette caractérisation appelle un changement de paradigme dans la manière de considérer les relations hôtes-virus qui va au-delà de la pathogénicité et aborde les autres types de relations trophiques. Un tel changement s'amorce progressivement pour le virome humain [10] et devient indispensable pour le virome de chaque espèce végétale [11, 12].

Ces changements vont aussi impacter les réglementations en vigueur pour la certification des plantes et pour la régulation du commerce mondial (quarantaine. . .). En effet, de nouveaux virus qui n'avaient jusqu'à présent pas été détectés par les techniques de diagnostic classiques pourraient être mis en évidence dans des plantes mères utilisées pour la multiplication intensive. Une telle découverte rétroactive soulèvera immanquablement de nombreuses questions concernant les mesures de protection à mettre en œuvre. Par ailleurs, la notion de plante certifiée indemne de virus (« *virus-free* ») pourrait devoir être revue de manière à prendre en compte les avancées scientifiques et le probable changement de paradigme, certains virus pouvant peut-être se révéler bénéfiques pour le phénotype de la plante.

En conclusion, le séquençage haut débit modifie d'ores et déjà la manière de détecter, de caractériser, de cataloguer et de considérer les viroses dans le domaine végétal comme dans les autres règnes du vivant. Son adoption pour le diagnostic de routine est à notre sens qu'une question de temps : les technologies de séquençage haut débit et de traitement des données continuent à évoluer favorablement, en particulier au niveau des coûts et de l'accessibilité. Dans le domaine du diagnostic phytopathologique, le rythme d'adoption dépendra bien sûr de l'objectif fixé (certification, contrôle aux frontières, etc.) et de l'espèce végétale ciblée (principalement valeur du matériel végétal et importance du risque phytosanitaire encouru). La validation de ces techniques sera plus complexe que la validation des techniques traditionnelles mais l'approbation récente par la Food and Drug Administration (FDA) de l'utilisation d'un séquenceur haut débit (Miseq-Illumina) pour le diagnostic représente une première étape très importante. Le séquençage haut débit pourra ainsi complémenter la PCR en temps réel pour améliorer le diagnostic. La profondeur de lecture permise par le séquençage NGS renforcera nos connaissances sur la variabilité des génomes viraux. Cela permettra notamment la définition de nouveaux jeux d'amorces PCR présentant une meilleure polyvalence et capables de détecter l'ensemble des isolats appartenant à une espèce virale. Nous sommes ainsi face à un saut

technologique majeur qui ouvre non seulement de nouvelles opportunités dans le diagnostic et la lutte contre les virus mais aussi de nouvelles voies de recherche sur les relations hôtes-virus dans tous les règnes du vivant.

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

Références

1. Boonham N, Kreuze J, Winter S, *et al*. Methods in virus diagnostics: From ELISA to next generation sequencing. *Vir Res* 2014 ; 186 : 20-31.

2. Barba M, Czosnek H, Hadidi A. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 2014 ; 6 : 106-36.

3. Massart S, Olmos A, Jijakli H, Candresse T. Current impact and future directions of high throughput sequencing in plant virus diagnostics. *Vir Res* 2014 ; 188 : 90-6.

4. Lecuit M, Eloit M. The diagnosis of infectious diseases by whole genome next generation sequencing: A new era is opening. *Front Cell Infect Microbiol* 2014 ; 4 : 25.

5. Cheval J, Sauvage V, Frangeul L, *et al*. Evaluation of high-throughput sequencing for identifying known and unknown viruses in biological samples. *J Clin Microbiol* 2011 ; 49 : 3268-75.

6. Hagen C, Frizzi A, Gabriels S, *et al*. Accurate and sensitive diagnosis of geminiviruses through enrichment, high-throughput sequencing and automated sequence identification. *Arch Virol* 2012 ; 157 : 907-15.

7. Eloit M. Heurs et malheurs du séquençage à haut débit en virologie. *Virologie* 2014 ; 18 : 11-21.

8. Massart S, Brostaux Y, Barbarossa L, *et al*. Inter-laboratory evaluation of a duplex RT-PCR method using crude extracts for the simultaneous detection of Prune dwarf virus and Prunus necrotic ringspot virus. *Eur J Plant Pathol* 2008 ; 122 : 539-47.

9. Gorbalenya AE. Taxonomic proposals based on metagenomic and genome-only studies. ICTV Newsletter. [Newsletter]. 2014 ; 11.

10. Lecuit M, Eloit M. The human virome: new tools and concepts. *Trends Microbiol* 2013 ; 21 : 510-5.

11. MacDiarmid R, Rodoni B, Melcher U, Ochoa-Corona F, Roossinck M. Biosecurity implications of new technology and discovery in plant virus research. *PLoS Pathog* 2013 ; 9 : e1003337.

12. Vayssier-Taussat M, Albina E, *et al*. Shifting the paradigm from pathogens to pathobiome new concepts in the light of meta-omics. *Front Cell Infect Microbiol* 2014 ; 4 : 29.